Scientific Report No. 12

Control Theory Group

ON MINIMUM FUEL AND ENERGY CONTROL OF SAMPLED-DATA CONTROL SYSTEMS

by

A. M. Revington and J. C. Hung

Supported by National Aeronautics and Space Administration

through Grant NsG-351

May 30, 1966

Department of Electrical Engineering

The University of Tennessee

ERRATA

Page	Location	Is	Should Be
11	line 5 from bottom	exisits	exists
24	line 2 from top	seqneuce	sequence
36	line 6 from bottom	symmetic	symmetric
45	line 12 from top	hyperplance	hyperplane
46	line 3 from bottom	...can, ...	...can; ...
52	Figure 17,	e^{1}, e^{2}	$e^{\gamma 1}$, $e^{\gamma 2}$
54	line after Eq. (2-95)	$\pm \underline{h}_3$	$+ \underline{h}_3$
66	Figure 25	Equations (2-98), (2-99), (2-101), (2-102), (2-103)	(2-109), (2-110), (2-112), (2-113), (2-114)
100	Equation (3-70)	$\delta^{et}\ \delta^{e}$	$\underline{\delta}^{et}\ \underline{\delta}^{e}$
145	Figure 55	u(1) = 0	u(1) = u(2)
247	Equation (A-41)	$e^{j \quad 1}$ etc.	$e^{j\gamma_1}$ etc.
248	Equation (A-43)	R_{ij} matrix notation	R_{ji} determinant notation
250	Equation (A-50)	$-e^{(\ )} - e^{(\ )}$	$-e^{(\ )} + e^{(\ )}$
255	Figure 83	Γ	Γ_3

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF FIGURES

FIGURE PAGE

FIGURE | PAGE

FIGURE | PAGE

CHAPTER I

INTRODUCTION

I. GENERAL DESCRIPTION

The general subject of this dissertation is the problem of nonlinear control of sampled-data systems. Only pulse-amplitude-modulated systems are considered; the control signal $u(t)$ being a piecewise constant function of time, t, which is allowed to change value only at periodic discrete instants of time T seconds apart; T is the sampling period. Such an input sequence is the output of a zero-order sample-hold device. The control signal is limited in magnitude by practical considerations. This type of control is known as saturating amplitude control. An example of such a control signal or sequence is given in Figure 1.

The dynamic system (plant) which is to be controlled is actuated by a controller. Figure 2 shows the configuration in block diagram form. The controller provides a control sequence, of the form of Figure 1, which is to take the system from an arbitrary initial state into (or close to) a desired state in a suitably prescribed manner. The controller receives information on the state of the plant only at discrete instants of time, T seconds apart. If the controller generates the input sequence on the basis of the initial state only, it will be called an open loop controller, and the corresponding input

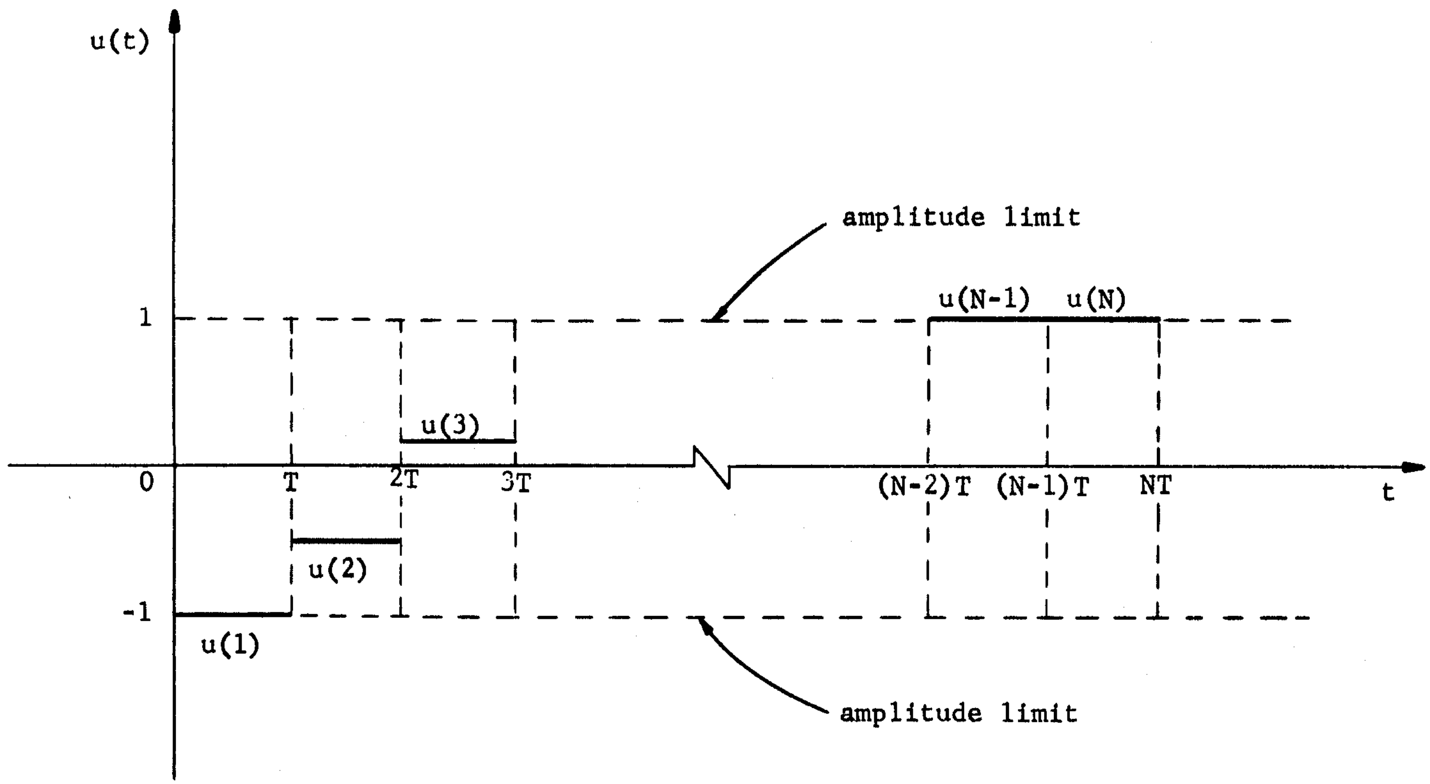


Figure 1. The form of the plant input.

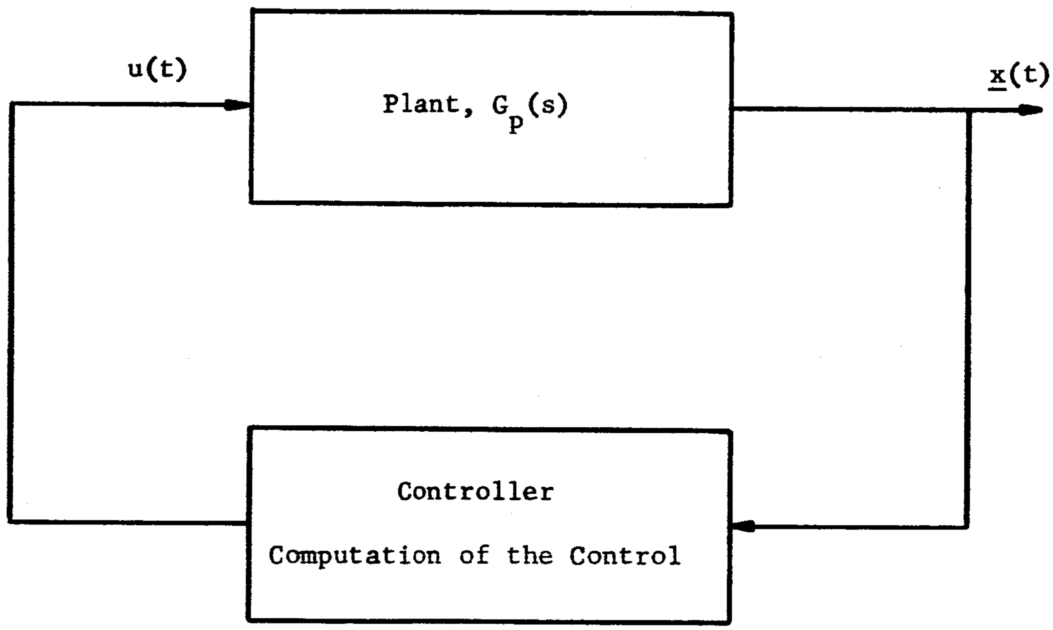


Figure 2. The system structure.

sequence an open loop control. If the input over the interval $kT \leqslant t < (k + 1)T$ is computed from the state at time kT, $k = 0, 1, \ldots,$ then the control will be called closed loop control or feedback control.

II. THE PLANT

The plants discussed in this dissertation are described by linear constant coefficient differential equations. The nonlinearity (saturation) is included in restrictions on the controller. It is assumed that the plant is controlled by a single input and that it is completely controllable (1, 2, 3, 4, 5, 6)*. Such a plant is commonly described by its transfer function, $G_p(s)$. Figure 3 shows the transfer function representing the plant in block diagram form. The order of the denominator polynomial, n, is the order of the plant; $-\lambda_i$ and $-z_i$ are respectively the poles and zeros of the plant. The plant may also be described by either a single n-th order differential equation or by n first order differential equations (4, 5, 6),

$$\frac{dx_i(t)}{dt} = \sum_{j=1}^{n} a_{ij}x_j(t) + d_i\, u(t), \quad i = 1, 2, \ldots, n, \tag{1-1}$$

where a_{ij} and d_i are constants. In matrix notation

$$\dot{\underline{x}}(t) = A\, \underline{x}(t) + \underline{d}\, u(t) \quad , \tag{1-2}$$

*Numbers in parentheses represent similarly numbered entries in the "List of References."

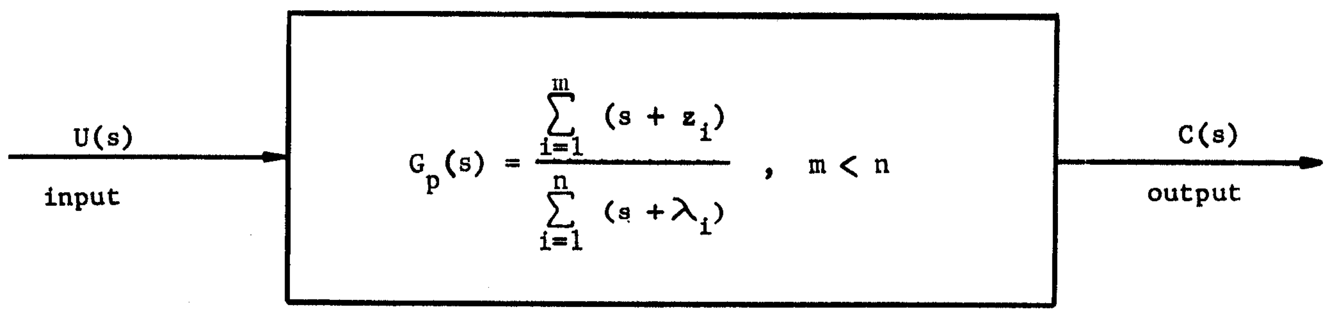


Figure 3. The transfer function of the plant.

where $(\cdot)$ denotes differentiation with respect to time. The vector $\underline{x}(t)$ is an n-vector (an n x 1 matrix), $\underline{d}$ is a constant n x 1 matrix and A is a constant n x n matrix. The vector $\underline{x}(t)$ defines a point in n-dimensional Euclidean Space, $\mathcal{X}$, with $x_1, \ldots, x_n$, the members of $\underline{x}$, forming a basis or coordinate system for the space. For a given control u(t) and initial point $x(t_o)$ the solution, see Appendix A, of Equation (1-2) describes a unique trajectory in $\mathcal{X}$. Given $\underline{x}(t_o)$ and u(t), $\underline{x}(t_o)$ is sufficient to describe the behaviour of the plant for any time $t \geq t_o$. For this reason $\underline{x}(t)$ is called the state of the plant and $\mathcal{X}$ is known as the state space. The elements of $\underline{x}$ $(x_1, \ldots, x_n)$ are called the state variables of the plant (4, 5, 6).

When u(t) is a piecewise constant input

$$u(t) = u(k), \quad \overline{k - 1}T \leq t < kT \quad , \quad k = 1, 2, \ldots, \tag{1-3}$$

the state of the plant at the discrete intervals of time kT, k = 0, 1, ..., is described by the following difference equation, derived in Appendix A:

$$\underline{x}(k + 1) = G(T)\ \underline{x}(k) + \underline{h}(T)\ u(k + 1) \tag{1-4}$$

where for convenience $\underline{x}(kT)$ is written as $\underline{x}(k)$. The matrix G(T) is n x n and is known as the transition matrix and h(T) is n x 1 and is called the forcing matrix. The plant is assumed to remain completely controllable in discrete form, see Appendix A. The state trajectory, moving under the influence of the control sequence u(k), k = 1, 2, ..., and the initial state $\underline{x}(0)$, are illustrated in Figure 4 for a second order system.

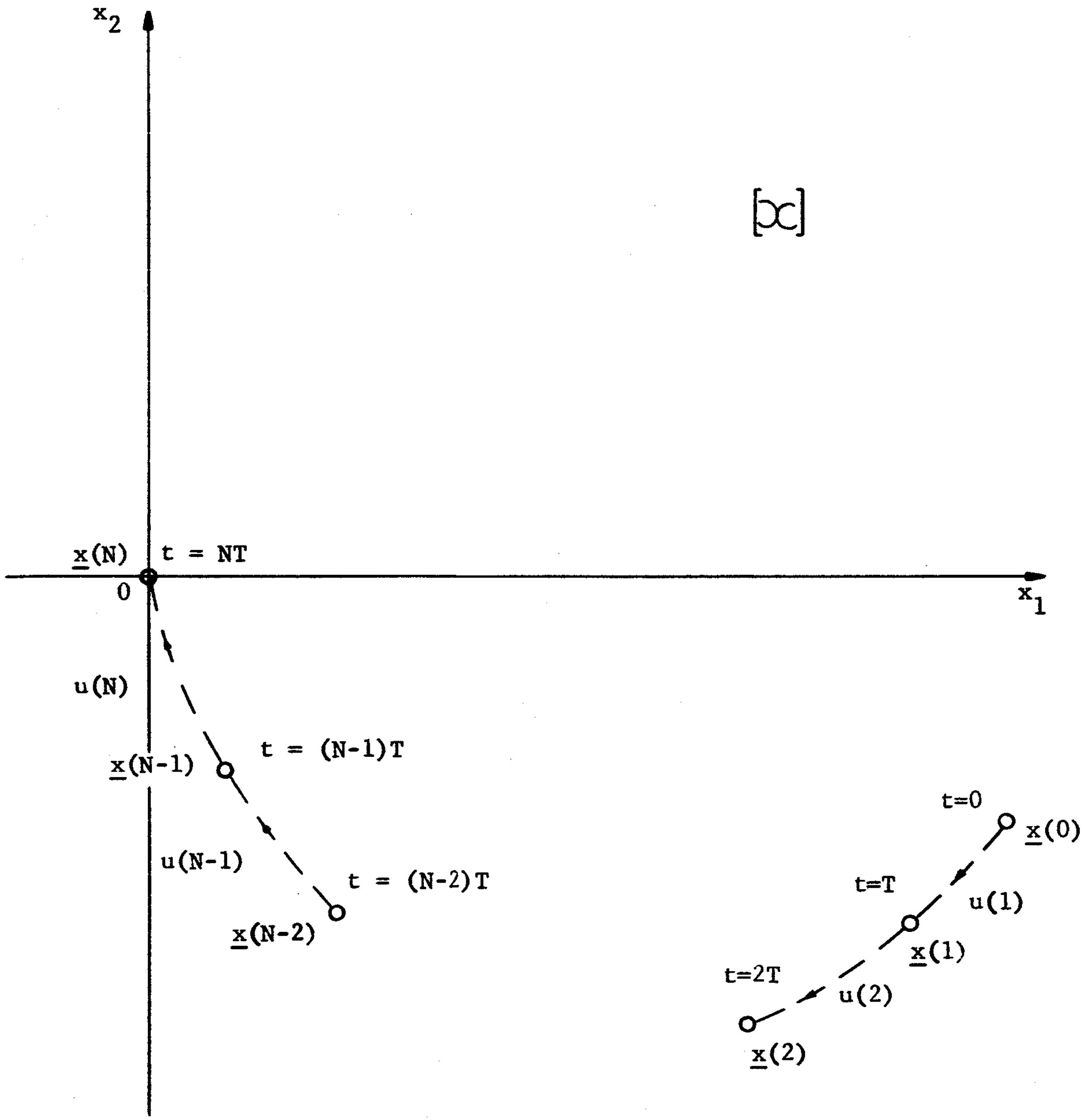


Figure 4. The state trajectory moving from $\underline{x}(0)$ to $\underline{x}(N) = 0$ under the influence of the input sequence $u(1), \ldots, u(N)$.

III. THE REGULATOR PROBLEM

The object of the control sequence u(k), k = 1, 2, ..., is to force the state of the plant from some arbitrary initial state $\underline{x}(0)$ to some desired final state in a suitable manner. For the regulator considered in this dissertation, it may be assumed that the origin of the state space is the desired final state. Upon reaching the origin the state will remain there if no further control signals are applied. Such a control is an example of deadbeat control (1, 7). The term "deadbeat control" has replaced the older Z-transform terminology "ripple-free error-free control" (8, 9, 10, 11, 12). The regulator problem may be described in terms of three factors: the length of time allowed for the regulatory process, the constraints on the process and the specification of the performance (subject to the constraints). These factors are discussed in turn.

The Time of Regulation

Let N be the total number of sampling periods allowed for the regulation. That is, after NT seconds the state of the plant has been forced from $\underline{x}(0)$ to $\underline{x}(N)$ and the regulation process is complete.

The Constraints

The desired final state is the origin of the state space; $\underline{x}(N) = 0$, and the control signal is limited in amplitude. Without loss of generality the amplitude is limited so that

$$\left|u(k)\right| \leqslant 1, \quad k = 1, 2, \ldots, N . \qquad (1\text{-}5)$$

The Performance Specification

The performance is usually considered optimum when some suitable cost function has been minimized. These cost functions are formulated to represent some physical desideratum. For example it has become customary to use the cost function

$$E = \sum_{k=1}^{N} \left[u(k)\right]^{2} \tag{1-6}$$

to represent the energy consumed by the control, and

$$F = \sum_{k=1}^{N} \left|u(k)\right| \tag{1-7}$$

to represent the fuel consumption. The main body of this dissertation will be limited to these two cost functions.

The three factors, time, constraints and cost function, which define the regulator problem cannot be specified independently. For example, minimizing the cost functions E and F has meaning only if some constraint like $\underline{x}(N) = 0$ is adjoined, and then only if there is more than one input sequence that can take $\underline{x}(0)$ to the origin. In Appendix A it is shown that if $\left|u(k)\right| \leqslant 1$, only a finite region of initial states, those in the set Γ_N, can be bought to the origin in N sampling periods or less. It will be assumed in formulating the regulator problem that N is always large enough for there to be a solution.

The regulator problems treated in this dissertation may be formalized as follows.

Minimum Energy Problem

Given $\underline{x}(0)$ in Γ_N find the input sequence $u(k)$, $k = 1, 2, \ldots, N$, such that

$$\underline{x}(N) = 0, \quad |u(k)| \leq 1, \quad \text{and} \quad E = \sum_{k=1}^{N} [u(k)]^2$$

is minimized.

Minimum Fuel Problem

Given $\underline{x}(0)$ in Γ_N find the input sequence $u(k)$, $k = 1, 2, \ldots, N$, such that

$$\underline{x}(N) = 0, \quad |u(k)| \leq 1, \quad \text{and} \quad F = \sum_{k=1}^{N} |u(k)|$$

is minimized.

IV. REVIEW OF THE REGULATOR PROBLEM

In order to place the particular problems chosen for discussion in perspective a brief review of several allied problems is given.

The Deadbeat Regulator

The term deadbeat control means that the state of the plant is forced to a desired state in a finite time, NT seconds, and remains there for $t \geq NT$. If the desired state is the origin of $\mathcal{X}$ then for $\underline{x}(t)$ to equal zero for $t \geq NT$, $u(t)$ must equal zero for $t \geq NT$.

This deadbeat condition, $\underline{x}(N) = 0$, is shown in Appendix A to be equivalent to the constraint

$$\underline{x}(0) = \sum_{j=1}^{N} \underline{r}_j \, u(j) \ , \tag{1-8}$$

where the $\underline{r}_j$ are the canonical vectors described in Appendix A. If the n x N matrix C is formed as

$$C = \left[\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N\right] \ , \tag{1-9}$$

and $\underline{u}$ = col. $\left[u(1), u(2), \ldots, u(N)\right]$, Equation (1-8) can be written as

$$C \, \underline{u} = \underline{x}(0) \ . \tag{1-10}$$

Equation (1-10) is the condition that the control $\underline{u}$ transfer $\underline{x}(0)$ to the origin in N sampling periods (13, 14, 15).

<u>The linear deadbeat regulator.</u> If the range of u(t) is not constrained by saturation the linear deadbeat problem is: Find $\underline{u}$ which minimizes a given cost function subject to $C \, \underline{u} = \underline{x}(0)$.

In time-optimal control it is desired to find the minimum N such that $C \, \underline{u} = \underline{x}(0)$. If $N < n$ there is no solution unless $\underline{x}(0)$ happens to be a linear combination of the N canonical vectors $\underline{r}_1$, $\underline{r}_2$, ..., $\underline{r}_N$. If N = n, C = R and for completely controllable plants the inverse of R exisits, see Appendix A. The unique time-optimal control sequence (7) is given by

$$\underline{u} = R^{-1} \, \underline{x}(0) \ . \tag{1-11}$$

When $N > n$, there is an infinite number of control sequences that satisfy Equation (1-10). The cost function which is to be minimized determines which of the control sequences may be used. Only one of these control sequences minimizes the energy cost function, E, while in some cases many of the allowable sequences may minimize the fuel cost function.

The generalized energy cost function is $\underline{u}^t S\, \underline{u}$ (2, 16, 17, 18). The transpose of a matrix is denoted by t. S is a positive definite N x N matrix. This cost function is of importance because S can be chosen to give a desired trajectory in the state space, and it is also easy to handle mathematically. The generalized energy problem is: minimize $\underline{u}^t S\, \underline{u}$ subject to $C\, \underline{u} = \underline{x}(0)$. The unique solution is

$$\underline{u} = S^{-1} C^t \left[C\, S^{-1} C^t \right]^{-1} \underline{x}(0) \tag{1-12}$$

and is developed in Chapter II. Bertram and Sarachik (16) solved this problem by using variational methods. They did not present the details. Kalman, Ho and Narendra (2) identified the problem with the generalized inverse (19, 20). Revington and Hung (18) and more recently, Yuji (21) reformulated and solved the problem using elementary differential calculus. This method is given in Chapter II. Cadzow (13) rediscovered Penrose's work (20), giving the solution Equation (1-12) for the case with S the identity matrix.

The fuel cost function, Equation (1-7), has not recieved much attention for discrete systems. Lee and Desoer (22) presented a formal solution to the linear fuel problem:

$$\text{minimize} \quad \sum_{k=1}^{N} \left| u(k) \right| \quad \text{subject to } C\, \underline{u} = \underline{x}(0).$$

The cost function is, in a mathematical sense, unsatisfactory because, as is shown in Chapter II, there may not be a unique solution to the problem. However, the practical importance of the fuel cost function requires that the problem be investigated.

The deadbeat regulator with saturation. In this case the range of u(t) is restricted as in Equation (1-5). The set of all (initial) states that can be brought to the origin in N or less sampling periods with saturating amplitude control is called Γ_N and is discussed in Appendix A. Kalman (23) defined the set and considered its properties; his work was extended by Desoer and Wing (14, 24, 25). Kurzweil (26) shows several of these sets for second order systems.

The deadbeat regulator with saturation is: Given $\underline{x}(0)$ in Γ_N, find a vector $\underline{u}$ which minimizes a given cost function subject to $C\,\underline{u} = \underline{x}(0)$ and $|u(k)| \leqslant 1$.

The time-optimal, minimum fuel and minimum energy problems are considerably complicated by the addition of the saturation constraint. As a solution to the time-optimal problem, Desoer and Wing in a planned series of papers (14, 24, 25) presented a method of constructing a switching surface which gives a feedback solution of practical importance for low order plants with real poles. Their methods were recently extended to cover state variable constraints (27). Tou (28) presented an open loop solution which works well in some cases. Torng (29) used linear programming concepts. Koepcke (30) used the digital computor to store information on the optimal input sequences so that real time feedback solutions could be obtained.

Ho (31) considered the "solution space", an N-dimensional Euclidean space with coordinates u(k), k = 1, 2, ..., N. This space is discussed briefly in Chapter III. With the input constraint, the admissible control region is a hypercube centered on the origin. The intersection of the hypercube with the (N-n)-dimensional hyperplane $C\,\underline{u} = \underline{x}(0)$ gives the feasible set of controls, whose members are control sequences that will take the initial state to the origin and satisfy the saturation constraint. In this formulation the time optimal problem consists of finding the smallest N such that there is an intersection between the hyperplane and hypercube, and then choosing one of the feasible controls. The minimum fuel and energy problems consist in finding from among the feasible controls one that minimizes the appropriate cost function. Viewing the problems in this light, Ho suggested that the fuel and energy problems were already solved since they were respectively simple linear and nonlinear programming problems. Torng (29) subsequently formulated the fuel problem and Kim (32) the energy problem in this manner. Such programming techniques (29, 30, 33, 34), however, are somewhat sterile in that they fail to give insight into the problems and intrinsically cannot suggest improvements to existing hardware. Furthermore, they cannot be used in a closed loop form with the present day requirements of real time solutions. Of course some control problems are so complex that general digital computer techniques must be utilized (35). The computer can be used to great advantage after all other avenues have been explored.

The most recent work on the energy problem is that of Stubberud and Swiger (36) who attempted to solve the energy problem in the solution space using intuition, functional analysis and set theory. Unfortunately, their conclusions, as shown in Chapter III, are not true in general.

Non-Deadbeat Control

This brief discussion of the regulator problem ought to mention non-deadbeat regulation. By removing the constraint $\underline{x}(N) = 0$; i.e, linear constraints of the form of Equation (1-10), and using cost functions of the form

$$\sum_{k=1}^{N} \underline{x}(k)^t \; P \; \underline{x}(k) + \mu \left[u(k)\right]^2 \quad , \tag{1-13}$$

with P a non-zero positive semidefinite matrix and $\mu \geqslant 0$, non-deadbeat regulators have been investigated. Kalman and Koepcke (1, 17, 37) and others (5, 38) treated the linear case and Deley and Franklin (39) considered the case with input saturation. Both solutions used dynamic programming, which is eminently suitable if N is large. If N is small, in the order of n, solutions in closed form are practical.

Perhaps the most interesting aspect of the non-deadbeat regulator is that it is very closely related to the discrete estimation problem (40, 41, 42).

V. SUMMARY OF THE WORK

The object of the dissertation is to study the problem of saturation in the minimum energy and minimum fuel deadbeat regulator, and to provide where possible, practical implementation of the optimal control in a feedback structure.

In Chapter II the theory of the linear energy and fuel problems is developed and used to consider in detail first and second order systems. Necessary and sufficient conditions for the uniqueness of the fuel solution are discussed. Chapter III discusses saturation in the minimum energy problem. It is shown that the open loop problem reduces to finding which members of the control sequence equal the saturation limit, ± 1. First order systems are solved completely as are certain second order systems. Chapter IV considers the corresponding minimum fuel problem. First order [illegible] systems are solved.

Chapter V gives suggestions as to how the optimal strategies may be implemented practically, in both open and closed loop forms. In certain cases very simple optimal and suboptimal strategies can be realized.

Appendix A provides the necessary background material for the dissertation and includes a discussion of the invariant vectors. Appendix B gives the derivation of some of the results used in Chapter III.

CHAPTER II

THE LINEAR ENERGY AND FUEL PROBLEMS

I. INTRODUCTION

Having formulated the minimum energy and minimum fuel problems in the canonical vector space, the derivation of the minimum energy equations is given. These equations are then extended to cover the generalized energy cost function discussed in Chapter I. A geometric approach to these equations results in a graphical method for estimating the minimum energy input sequence, which is particularly useful for second order systems.

While the minimum energy problem is solved by differential calculus, the minimum fuel problem is approached by considering a set, $S_N(f)$. An

before its general properties are presented. The optimum input sequence for first order systems can be solved without the explicit use of this set, but the general properties of the set do provide the comfort of rigor for second and higher order systems. In discussing first and second order systems, it is shown that the input sequence is not necessarily unique. Theorem 1 gives the necessary and sufficient conditions for the uniqueness of the minimum fuel input sequence. Second order plant pole configurations, for which initial conditions occur with non-unique sequences, are investigated.

The chapter closes with a detailed example of both the minimum energy and minimum fuel problem.

II. FORMULATION IN $\mathcal{C}$-SPACE

The linear deadbeat regulator with minimum energy is equivalent to the problem,

$$\text{minimize } E = \sum_{j=1}^{N} \left[u(j)\right]^2 \quad \text{subject to} \quad \sum_{j=1}^{N} u(j)\,\underline{r}_j = \underline{x}(0) \;. \tag{2-1}$$

The corresponding minimum fuel problem is,

$$\text{minimize } F = \sum_{j=1}^{N} \left|u(j)\right| \quad \text{subject to} \quad \sum_{j=1}^{N} u(j)\,\underline{r}_j = \underline{x}(0) \;. \tag{2-2}$$

If the input sequence is arranged as the N x 1 column vector, $N > n$,

$$\underline{u} = \text{col.}\left[u(1),\ u(2),\ \ldots,\ u(N)\right] \;, \tag{2-3}$$

and the canonical vectors are arranged in matrix form as

$$C = \left[\underline{r}_1,\ \underline{r}_2,\ \ldots,\ \underline{r}_N\right] \;, \tag{2-4}$$

where C is an n x N matrix, problems (2-1) and (2-2) become respectively:

$$\text{minimize } E = \underline{u}^t\,\underline{u} \quad \text{subject to} \quad C\underline{u} = \underline{x}(0) \;; \tag{2-5}$$

$$\text{minimize } F = \sum_{j=1}^{N} \left|u(j)\right| \quad \text{subject to} \quad C\underline{u} = \underline{x}(0) \;. \tag{2-6}$$

The transformation of these problems to the canonical vector space, $\mathcal{C}$, can be considered in the following manner. Let C be partitioned as

$$C = \left[R,\ Q\right] \quad , \tag{2-7}$$

where

$$R = \left[\underline{r}_1,\ \underline{r}_2,\ \ldots,\ \underline{r}_n\right] \quad , \tag{2-8}$$

$$Q = \left[\underline{r}_{n+1},\ \underline{r}_{n+2},\ \ldots,\ \underline{r}_N\right] \quad . \tag{2-9}$$

Let $\underline{u}$ be partitioned into

$$\underline{a} = \text{col.}\left[u(1),\ u(2),\ \ldots,\ u(n)\right] \quad , \tag{2-10}$$

$$\underline{b} = \text{col.}\left[u(n+1),\ u(n+2),\ \ldots,\ u(N)\right] \quad . \tag{2-11}$$

The deadbeat constraint, $C\underline{u} = \underline{x}(0)$, becomes

$$\underline{x}(0) = \left[R,\ Q\right]\begin{bmatrix}\underline{a}\\ \underline{b}\end{bmatrix} \quad , \tag{2-12}$$

$$= R\underline{a} + Q\underline{b} \quad . \tag{2-13}$$

Premultiplying Equation (2-13) by R^{-1}, and defining

$$\underline{c} = R^{-1}\ \underline{x}(0) \quad , \tag{2-14}$$

$$H = R^{-1}\ Q \quad , \tag{2-15}$$

gives

$$\underline{c} = \underline{a} + H\underline{b} = \sum_{j=1}^{N} u(j)\ \underline{h}_j \quad , \tag{2-16}$$

where the "invariant vectors"* are

$$\underline{h}_j = R^{-1}\ \underline{r}_j \quad , \qquad j = 1,\ 2,\ \ldots,\ N \quad .$$

---

*The $\underline{h}_j$ vectors so defined are different from those used in references 15, 18, 45 and 48.

Thus, the $n x(N - n)$ matrix H is composed of the last N - n invariant vectors (Appendix A):

$$H = \left[\underline{h}_{n+1}, \underline{h}_{n+2}, \ldots, \underline{h}_N\right] \quad . \tag{2-17}$$

Equation (2-16) is the deadbeat constraint in $\mathcal{C}$-space. In future the state space, $\mathcal{X}$, will be referred to only occasionally.

III. THE MINIMUM ENERGY PROBLEM

Problem (2-5) becomes,

$$\text{minimize} \quad E = \underline{a}^t\underline{a} + \underline{b}^t\underline{b} \quad \text{subject to } \underline{c} = \underline{a} + H\underline{b} \quad . \tag{2-18}$$

The solution is as follows:

$$E = \underline{a}^t\underline{a} + \underline{b}^t\underline{b} \quad , \tag{2-19}$$

$$= \left[\underline{c} - H\underline{b}\right]^t\left[\underline{c} - H\underline{b}\right] + \underline{b}^t\underline{b} \quad , \tag{2-20}$$

$$= \underline{c}^t\underline{c} - 2\underline{b}^t H^t\underline{c} + \underline{b}^t\left[I + H^tH\right]\underline{b} \quad . \tag{2-21}$$

Taking the gradient of E with respect to $\underline{b}$ (5, page 45; 43, page 45) gives

$$\nabla_{\underline{b}} E = 2\left[I + H^tH\right]\underline{b} - 2H^t\underline{c} \quad . \tag{2-22}$$

Setting $\nabla_{\underline{b}}E = 0$ gives the condition for E to be a minimum. Therefore, the optimum $\underline{b}$, $\underline{b}^o$, is given by

$$\left[I + H^tH\right]\underline{b}^o = H^t\underline{c} \quad . \tag{2-23}$$

From Equations (2-16) and (2-23) with the optimal $\underline{a}$ given by $\underline{a}^o$, there results

$$\underline{b}^o + H^tH\underline{b}^o = H^t\underline{a}^o + H^tH\underline{b}^o \quad . \tag{2-24}$$

Therefore, the condition for minimum energy in the linear deadbeat regulator is simply (15, page 8; 21, page 836)

$$\underline{b}^o = H^t \underline{a}^o \quad . \tag{2-25}$$

From Equation (2-16) and (2-25),

$$\left[I + HH^t\right] \underline{a}^o = \underline{c} \quad . \tag{2-26}$$

Since the system is completely controllable any n, and only n, of the invariant vectors are linearly independent. This means that H is of maximal rank. It follows that the n x n matrix in Equation (2-26) can be inverted and is in fact positive definite (18, page 13). Then

$$\underline{a}^o = \left[I + HH^t\right]^{-1} \underline{c} \tag{2-27}$$

and from Equation (2-25),

$$\underline{b}^o = H^t \left[I + HH^t\right]^{-1} \underline{c} \quad . \tag{2-28}$$

Alternatively, from Equation (2-23),

$$\underline{b}^o = \left[I + H^t H\right]^{-1} H^t \underline{c} \quad . \tag{2-29}$$

Greville (44) and Cadzow (13) obtained the solution to the same problem in a different form which can readily be obtained from Equations (2-27) and (2-28). Equation (2-27) gives, with Equation (2-7),

$$\underline{a}^o = \left[I + R^{-1} QQ^t R^{-1t}\right]^{-1} R^{-1} \underline{x}(0) \quad , \tag{2-30}$$

$$= \left[R^{-1}(RR^t + QQ^t)R^{-1t}\right]^{-1} R^{-1} \underline{x}(0) \quad , \tag{2-31}$$

$$= R^t \left[RR^t + QQ^t\right]^{-1} \underline{x}(0) \quad . \tag{2-32}$$

Equation (2-28) gives

$$\underline{b}^o = Q^t R^{-1t}\left[RR^t + QQ^t\right]^{-1} \underline{x}(0) \quad , \tag{2-33}$$

$$= Q^t\left[RR^t + QQ^t\right]^{-1} \underline{x}(0) \quad . \tag{2-34}$$

Combining Equations (2-32) and (2-34) gives

$$\underline{u}^o = \begin{bmatrix} \underline{a}^o \\ \underline{b}^o \end{bmatrix} = \begin{bmatrix} R^t \\ Q^t \end{bmatrix} \left[RR^t + QQ^t\right]^{-1} \underline{x}(0) \quad , \tag{2-35}$$

$$= \left[R, Q\right]^t \left[(R, Q)(R, Q)^t\right]^{-1} \underline{x}(0) \quad , \tag{2-36}$$

$$= C^t\left[CC^t\right]^{-1} \underline{x}(0) \quad . \tag{2-37}$$

This solution is certainly more compact than the solution given by Equations (2-27) and (2-28). However, the solution in $\mathcal{C}$ -space is much more useful because it is independent of the state space coordinate system. Furthermore, Equation (2-25) allows very useful geometric pictures to be used in considering both the linear and saturating minimum energy problems.

The minimum cost, E^o, in $\mathcal{X}$ -space is

$$E^o = \underline{u}^{ot} \underline{u}^o = x(0)^t\left[CC^t\right]^{-1} \underline{x}(0) \quad , \tag{2-38}$$

and in $\mathcal{C}$ -space, is

$$E^o = \underline{a}^{ot} \underline{a}^o + \underline{b}^{ot} \underline{b}^o = \underline{c}^t\left[I + HH^t\right]^{-1} \underline{c} = \underline{c}^t \underline{a}^o \quad . \tag{2-39}$$

<u>The Generalized Energy Cost Function</u>

Consider now the cost function $\underline{u}^t S \underline{u}$, where S is an N x N positive definite matrix. The solution given in Chapter I, Equation

(1-13), can be obtained from Equation (2-37) in a straightforward manner.

It is always possible to obtain an invertible N x N matrix, D, so that S may be written as

$$S = D^t D \quad , \tag{2-40}$$

giving

$$\underline{u}^t S \underline{u} = \underline{u}^t D^t D \underline{u} \quad . \tag{2-41}$$

Define

$$\underline{v} = D \underline{u} \quad . \tag{2-42}$$

The generalized energy problem, minimize $\underline{u}^t S\underline{u}$ subject to $C\underline{u} = \underline{x}(0)$, is therefore equivalent to the problem, minimize $\underline{v}^t\underline{v}$ subject to $CD^{-1}\underline{v} = \underline{x}(0)$. The solution to this latter problem is obtained from Equation (2-37), and is

$$\underline{v}^o = (CD^{-1})^t \left[CD^{-1}(CD^{-1})^t\right]^{-1} \underline{x}(0) \quad , \tag{2-43}$$

$$= \left[D^t\right]^{-1} C^t \left[CD^{-1}\left[D^{-1}\right]^t C^t\right]^{-1} \underline{x}(0) \quad . \tag{2-44}$$

Using Equations (2-40) and (2-42),

$$\underline{u}^o = S^{-1} C^t \left[CS^{-1}C^t\right]^{-1} \underline{x}(0) \quad . \tag{2-45}$$

This is the solution to the generalized energy problem.

The cost function $\underline{u}^t S\underline{u}$ is very practical. By a suitable choice of S, factors such as the risetime and overshoot can be made to meet practical specifications while maintaining the deadbeat response (18, page 10). The dynamic programming approach (17), which has similar advantages, is not deadbeat.

Geometric Interpretation and Solution of the Energy Problem

A graphical method (45) of finding the input seqneuce for a given initial state $\underline{c}$ will be described for second order systems. The concepts are equally applicable to higher order systems.

The set of all initial states that give $u^o(j)$ = constant is seen to be a hyperplane. The equation of the hyperplane may be found directly from Equation (2-27) or Equation (2-28). If the lines $u^o(j) = 0$ and $u^o(j) = 1$ are drawn in $\mathcal{C}$-space, the j-th control can be found for a given initial state $\underline{c}$ by linear interpolation or extrapolation. A method of obtaining these lines without solving the equations directly can be approached through the use of an auxiliary space, $\mathcal{A}^o$-space.

Consider the n inputs $u^o(1)$, $u^o(2)$, ..., $u^o(n)$ as the coordinates of an n-dimensional Euclidean space, $\mathcal{A}^o$-space. The transformation between $\mathcal{C}$ and $\mathcal{A}^o$ is given by Equation (2-27),

$$\underline{a}^o = \left[I + HH^t\right]^{-1} \underline{c} \quad .$$

For second order systems (n = 2), $\underline{h}_j^t\, \underline{a}^o = u^o(j)$, j = 1, 2, ..., N are lines and are normal to the corresponding vector $\underline{h}_j$. For $u(j) = 0$ the line passes through the origin. For $u^o(j) > 0$ the line moves in the direction of $+\underline{h}_j$ and conversely for $u^o(j) < 0$. When $u^o(j) = 1$ the line intersects $\underline{h}_j$ at a distance $1/\ell_j$ from the origin, where

$$\ell_j = (\underline{h}_j^t\, \underline{h}_j)^{1/2} \quad , \tag{2-46}$$

is the length of $\underline{h}_j$. Figure 5 shows the configuration. After plotting these lines for j = 1, 2, ..., N, the structure of the optimal control sequence may be investigated.

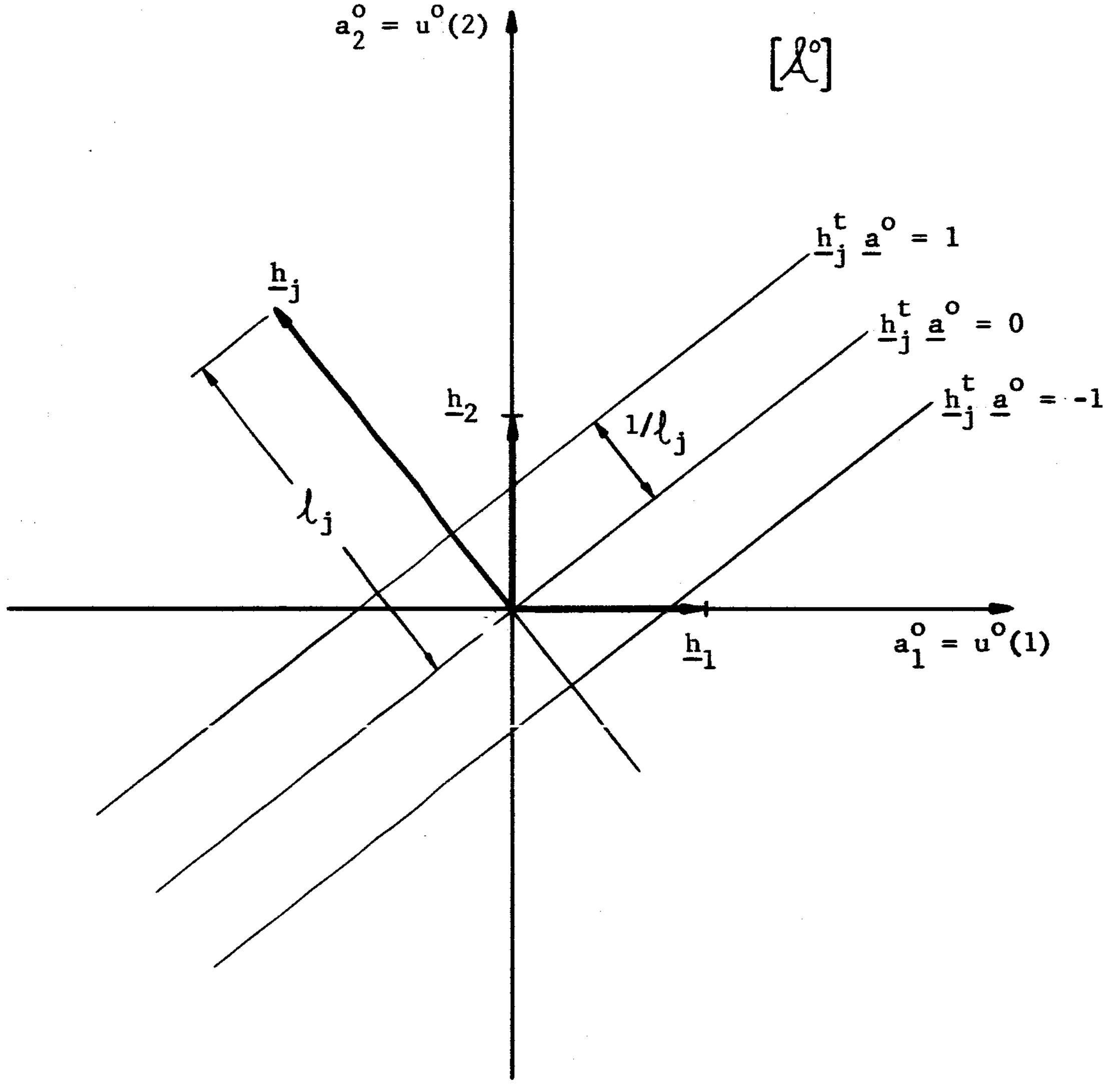


Figure 5. The invariant vector $\underline{h}_j$ in $\mathcal{L}^o$-space.

It remains to link the lines in $\mathcal{A}^o$ to the initial states in $\mathcal{C}$. From Equation (2-27), or directly from Equation (2-16),

$$\underline{c} = \underline{a}^o + H\underline{b}^o = \sum_{j=1}^{N} u^o(j)\ \underline{h}_j \quad . \tag{2-47}$$

By considering two convenient points on the line $\underline{h}_j^t\ \underline{a}^o = 1$, the corresponding $\underline{a}^o$ and $\underline{b}^o$ can be estimated. The corresponding $\underline{c}$ can be constructed in $\mathcal{C}$ by adding together the vectors $u^o(j)\ \underline{h}_j$.

In actual practice there is no need to draw two diagrams. The lines in $\mathcal{A}^o$ can be drawn directly in $\mathcal{C}$ by imagining that the coordinates $u^o(1), \ldots, u^o(n)$ replace the coordinates $c_1, \ldots, c_n$. The construction lines may be ignored once $u^o(j) = 1$ has been drawn in $\mathcal{C}$.

This technique is of course practical for first and second order systems only, but the principle holds for any n. The control sequence can be estimated quite accurately if N is not too large, but even if the technique cannot be used to obtain the control sequence exactly, a rough idea of the structure of the control can in itself be very useful. The technique is used to advantage in Chapter III.

IV. THE MINIMUM FUEL PROBLEM

The minimum fuel problem, Problem (2-6), is

$$\text{minimize } F = \sum_{j=1}^{N} |u(j)| \quad \text{subject to } \underline{c} = \sum_{j=1}^{N} u(j)\ \underline{h}_j \ . \tag{2-48}$$

Introductory Discussion

Since the fuel cost function F cannot be handled by conventional differential calculus, the general properties of the minimum fuel sequence are introduced by considering a second order system with a settling time of three sampling periods; i.e, $n = 2$ and $N = 3$. The most important characteristics of the input sequence can be demonstrated with the plant

$$G_p(s) = \frac{1}{s^2} \quad . \tag{2-49}$$

Figure 6 shows the invariant vectors $\underline{h}_1$, $\underline{h}_2$, and $\underline{h}_3$ for this plant. The characteristics of the input sequence will be examined by consideration of three initial states.

1. Consider the initial state $\underline{c} = \underline{h}_1$. One possible input sequence, satisfying the constraint in Problem (2-48), is clearly

$$u(1) = 1, \quad u(2) = 0, \quad u(3) = 0, \tag{2-50}$$

and the fuel cost is $F = 1$. Is there another input sequence that satisfies the constraint and costs less fuel? In an attempt to reduce the cost, u(1) must be reduced. Suppose u(1) is reduced to α: $u(1) = \alpha$, $0 < \alpha < 1$. Then since

$$\sum_{j=1}^{3} u(j)\, \underline{h}_j = \underline{h}_1 = \underline{c} \quad , \tag{2-51}$$

u(2) and u(3) must satisfy

$$\underline{h}_1(1 - \alpha) = u(2)\underline{h}_2 + u(3)\underline{h}_3 \quad . \tag{2-52}$$

Therefore,

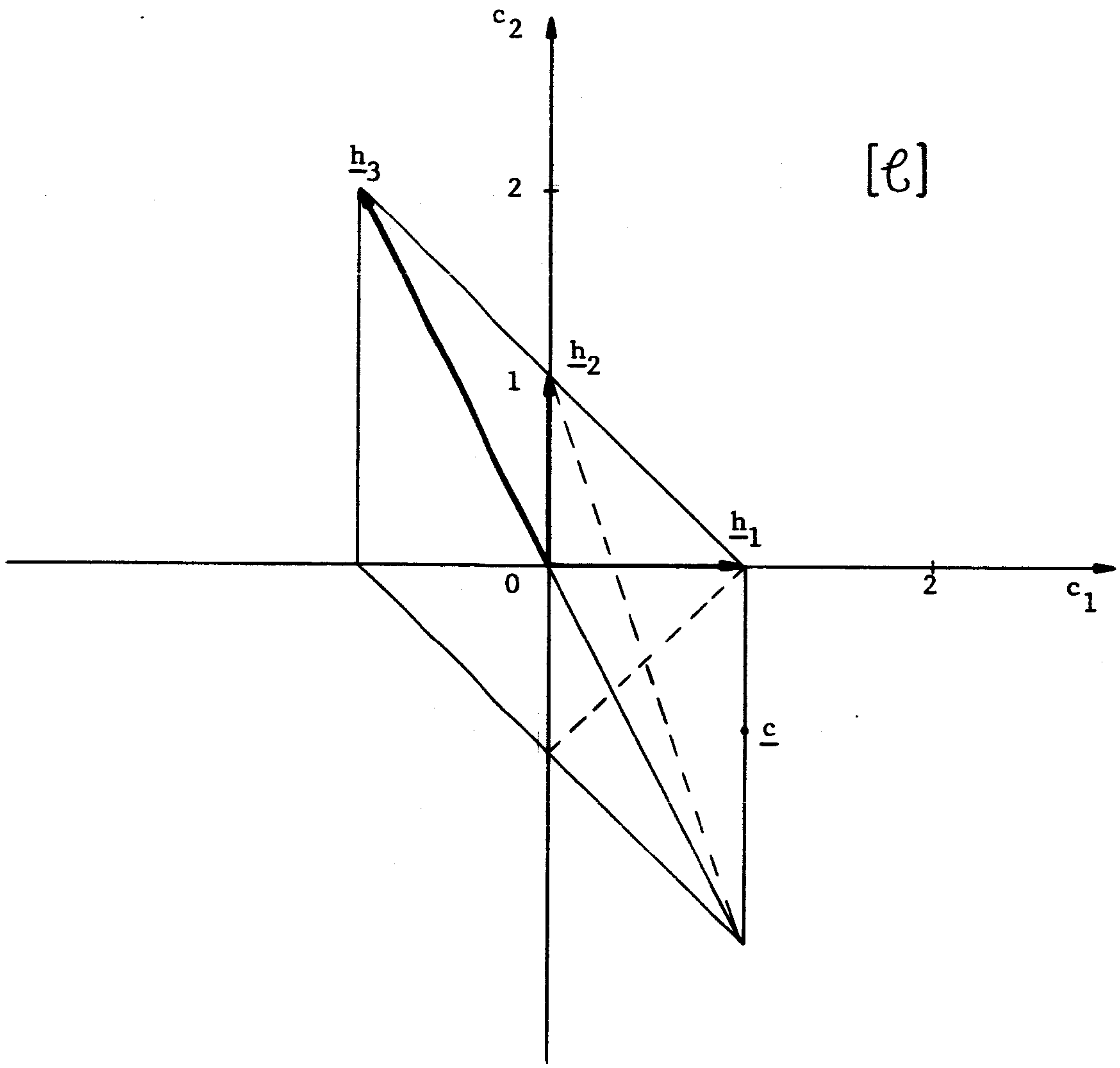


Figure 6. The invariant vectors $\underline{h}_1$, $\underline{h}_2$ and $\underline{h}_3$ for the plant $1/s^2$.

$$\begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} u(2) \\ u(3) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} (1 - \alpha) \quad , \tag{2-53}$$

which gives

$$u(2) = 2 - 2\alpha, \quad u(3) = -1 + \alpha \quad . \tag{2-54}$$

The fuel cost is therefore

$$\begin{aligned} F &= |u(1)| + |u(2)| + |u(3)| \\ &= 3 - 2\alpha \quad . \end{aligned} \tag{2-55}$$

For $1 > \alpha > 0$, $F > 1$. This means that the input sequence, Equation (2-50), is a unique optimum input sequence for $\underline{c} = \underline{h}_1$.

2. Consider next the initial state $\underline{c} = \underline{h}_2$. Possible input sequences are seen to be

$$\begin{aligned} &u(1) = 0, \quad u(2) = 1, \quad u(3) = 0, \\ &u(1) = \tfrac{1}{2}, \quad u(2) = 0, \quad u(3) = \tfrac{1}{2}, \\ &u(1) = \tfrac{1}{3}, \quad u(2) = \tfrac{1}{3}, \quad u(3) = \tfrac{1}{3}, \\ &u(1) = \tfrac{3}{8}, \quad u(2) = \tfrac{1}{4}, \quad u(3) = \tfrac{3}{8}, \end{aligned} \tag{2-56}$$

and again $F = 1$. Letting $u(2) = \alpha$, $0 \leqslant \alpha \leqslant 1$ in an attempt to find a sequence with less fuel consumption, the deadbeat constraint is,

$$\underline{h}_2 = u(1)\, \underline{h}_1 + \alpha\, \underline{h}_2 + u(3)\, \underline{h}_3 \quad . \tag{2-57}$$

Therefore, on solving for u(1) and u(3),

$$u(1) = \frac{1}{2} - \frac{\alpha}{2} \quad , \quad u(3) = \frac{1}{2} - \frac{\alpha}{2} \quad , \tag{2-58}$$

and

$$F = \frac{1}{2} - \frac{\alpha}{2} + \frac{1}{2} - \frac{\alpha}{2} + \alpha = 1 \quad . \tag{2-59}$$

Therefore F is independent of α, $0 \leqslant \alpha \leqslant 1$, and there is consequently no way of obtaining an input sequence with $F < 1$. The non-unique minimum fuel solution is therefore

$$u(1) = \frac{1-\alpha}{2}, \quad u(2) = \alpha, \quad u(3) = \frac{1-\alpha}{2}. \tag{2-60}$$

3. As a final example consider an initial state $\underline{c}$ on the line joining $\underline{h}_1$ and $-\underline{h}_3$. Such a state can be described by

$$\underline{c} = \mu_1 \underline{h}_1 - \mu_2 \underline{h}_3, \quad \mu_1, \mu_2 \geqslant 0, \quad \mu_1 + \mu_2 = 1. \tag{2-61}$$

One possible input sequence is

$$u(1) = \mu_1, \quad u(2) = -\mu_2, \quad u(3) = 0, \tag{2-62}$$

and $F = 1$. Is it possible to find another input sequence giving a smaller fuel cost? Let $\underline{c}$, shown in Figure 6, page 28, be a typical initial state given by Equation (2-61). Now consider states on the dashed line joining $\underline{h}_2$ and $-\underline{h}_3$. With $u(1) = 0$, such states can be taken to the origin with a minimum fuel cost $F = 1$. Similarly, with $u(3) = 0$, states lying on the dashed line joining $\underline{h}_1$ and $-\underline{h}_2$ can be taken to the origin with minimum fuel cost $F = 1$. Now the initial state $\underline{c}$ may be represented by linearly combining either $\underline{h}_1$ and $\underline{h}_2$, or $\underline{h}_2$ and $\underline{h}_3$, or $\underline{h}_1$ and $\underline{h}_3$, or finally, $\underline{h}_1$, $\underline{h}_2$ and $\underline{h}_3$. Considering Figure 6, page 28, the combination of $\underline{h}_1$ and $\underline{h}_2$ would require a fuel cost exceeding $F = 1$, since $\underline{c}$ lies beyond the dashed line joining $\underline{h}_1$ and $-\underline{h}_2$. Similarly, since $\underline{c}$ lies beyond the dashed line joining $\underline{h}_2$ and $-\underline{h}_3$, the second combination, $\underline{h}_2$ and $\underline{h}_3$, would also require $F > 1$. The third combination gives the input sequence of Equation (2-62), which makes

$F = 1$. It therefore remains to see if any reduction in fuel cost can be obtained if all three invariant vectors ($\underline{h}_1$, $\underline{h}_2$ and $\underline{h}_3$) are used to represent $\underline{c}$. Suppose u(2) is fixed at some value, α. The representation of $\underline{c}$ is therefore

$$\underline{c} - \alpha\,\underline{h}_2 = u(1)\,\underline{h}_1 + u(3)\,\underline{h}_3 \quad . \tag{2-63}$$

If $\alpha = 0$, Equation (2-61) gives $u(1) = \mu_1$ and $u(2) = -\mu_2$, with $F = 1$. If α is increased from zero, $\underline{c} - \alpha\,\underline{h}_2$ moves from $\underline{c}$ along a straight line at $\underline{c}$ and passing through $-\underline{h}_3$. Similarly if α is decreased from zero, $\underline{c} - \alpha\,\underline{h}_2$ moves from $\underline{c}$ and passes through $\underline{h}_1$. With $\underline{c}$ given by Equation (2-63) therefore, the fuel cost is never less than $F = 1 + |\alpha|$. Therefore, the optimum input sequence, for an initial state on the line joining $\underline{h}_1$ and $-\underline{h}_3$, is uniquely given by Equation (2-62).

The results on these initial states can be combined and extended. Before proceeding, however, it is necessary to know what is meant by a "cone." A cone is defined as follows (46, page 219): A cone is a set of points with the following property: if $\underline{c}$ is in the set, so is $\mu\underline{c}$ for all $\mu \geqslant 0$.

Consider the set of points, $L(\pm i, \pm j)$ on the line joining $\pm\,\underline{h}_i$ to $\pm\,\underline{h}_j$, $i \neq j$. The dashed lines in Figure 7 show $L(i,j)$, $L(j,-i)$, $L(-i,-j)$ and $L(-j,i)$ for a typical pair of invariant vectors $\underline{h}_i$ and $\underline{h}_j$. The cone "generated" by the set of points $L(\pm i, \pm j)$ is defined to be the set

$$C(\pm i, \pm j) = \left\{ \underline{y} \,\middle|\, \underline{y} = \mu\underline{c}, \text{ all } \mu \geqslant 0 \text{ and all } \underline{c} \text{ in } L(\pm i, \pm j) \right\}. \tag{2-64}$$

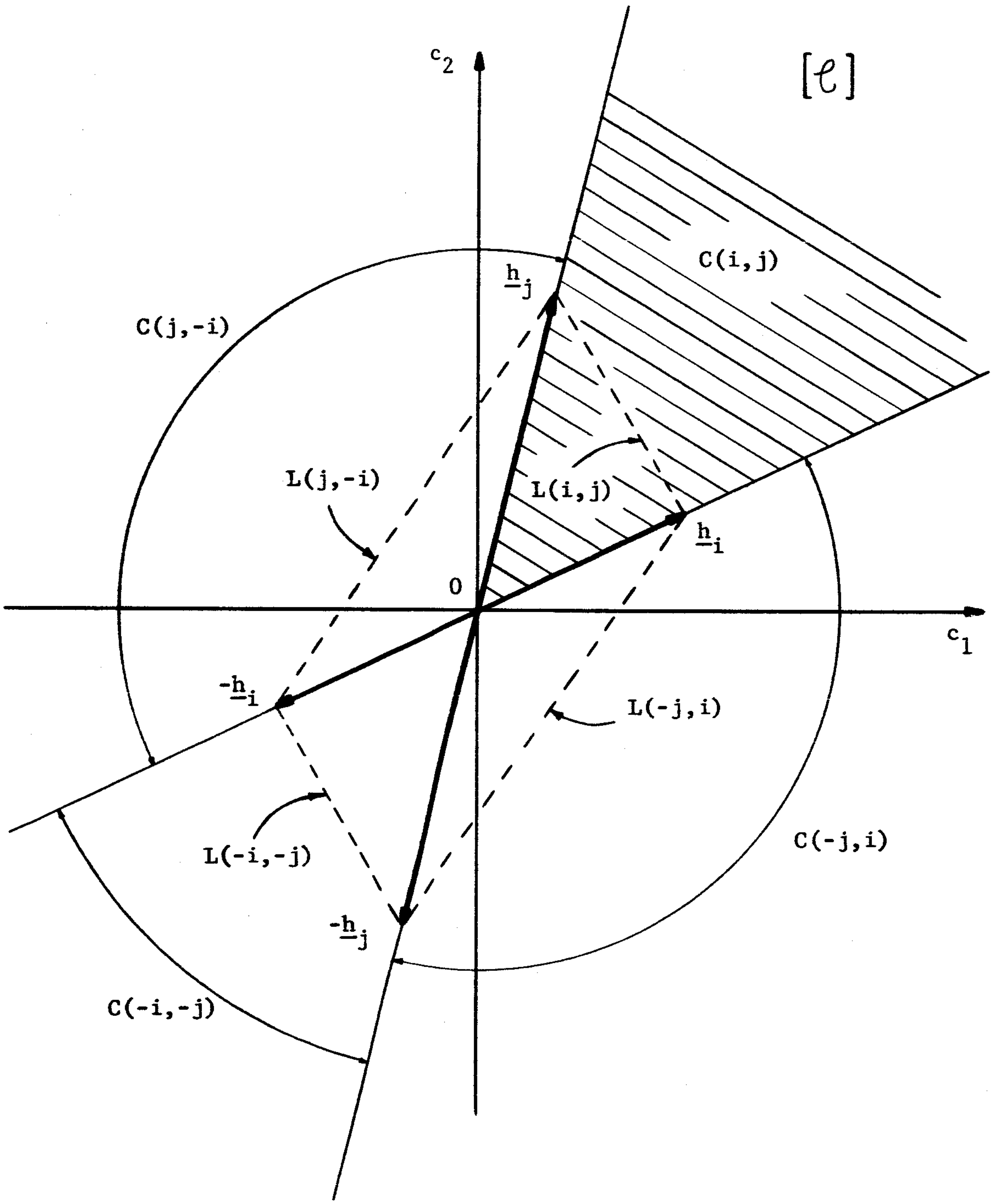


Figure 7. The cone C($\pm$i,$\pm$j) generated by the line L($\pm$i,$\pm$j).

For example, the cone C(i,j) is the cross hatched region indicated in Figure 7. Further, note that, taken together, the four cones C(i,j), C(j,-i), C(-i,-j) and C(-j,i) cover the entire $\mathcal{C}$ -space.

Now consider the example again. The "smallest" convex set which contains the points $\pm f\underline{h}_1$, $\pm f\underline{h}_2$, $\pm f\underline{h}_3$, where $f > 0$ is called $S_3(f)$ after the notation of Lee and Desoer(22). The boundary of $S_3(f)$ is called $\partial S_3(f)$. Figure 8 shows the set $S_3(f)$ and its boundary $\partial S_3(f)$. Figure 9 shows $\mathcal{C}$ -space divided into six regions by the cones $C(\pm i, \pm j)$, i, j = 1, 2, 3.

Although, for the sake of simplicity, the initial states discussed above were assumed to lie on $\partial S_3(1)$, the characteristics of the optimal input sequence when the initial state lies on $\partial S_3(f)$ are identical except that the input members are f times greater and F = f. Therefore, when $\underline{c}$ lies on $\partial S_3(f)$, the following observations can be made:

1. If $\underline{c}$ is in C(1,2) or C(2,3), C(-1,-2) or C(-2,-3), the optimal control sequence is not unique. If $\underline{c}$ is in C(-3,1) or C(3,-1), the optimal control sequence is unique.

2. The optimum fuel cost is F = f.

3. An optimum input sequence can always be found by using an input sequence with only one or two non-zero members. For example, if $\underline{c}$ lies in C(i,j), $i,j = \pm 1, \pm 2, \pm 3$, the input sequence can be obtained by representing $\underline{c}$ as

$$\underline{c} = \mu_1 \underline{h}_i + \mu_2 \underline{h}_j, \quad i,j = 1, 2, 3 \quad , \tag{2-65}$$

with

[ℓ]

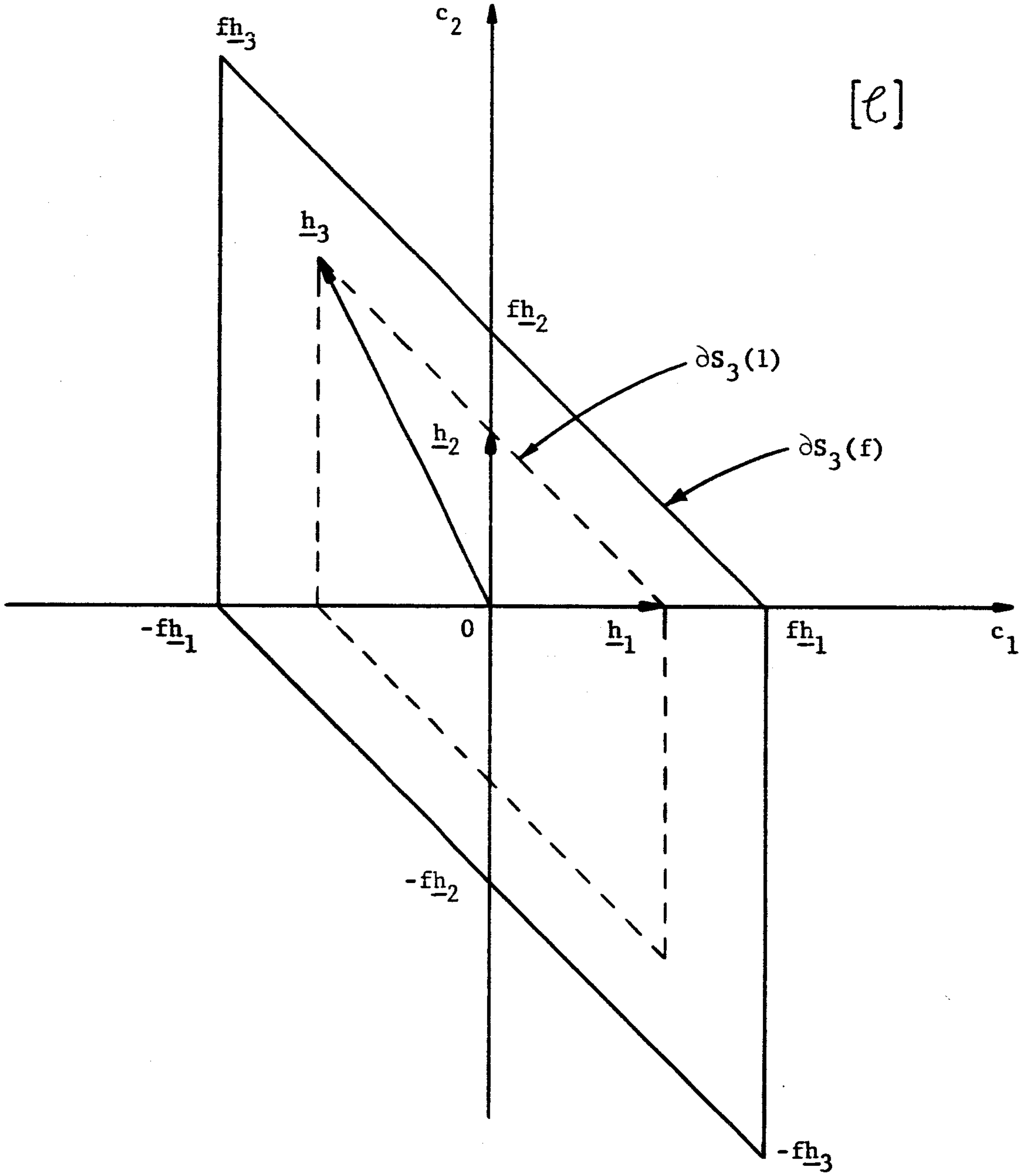


Figure 8. The boundaries $\partial S_3(1)$ and $\partial S_3(f)$ for the plant $1/s^2$.

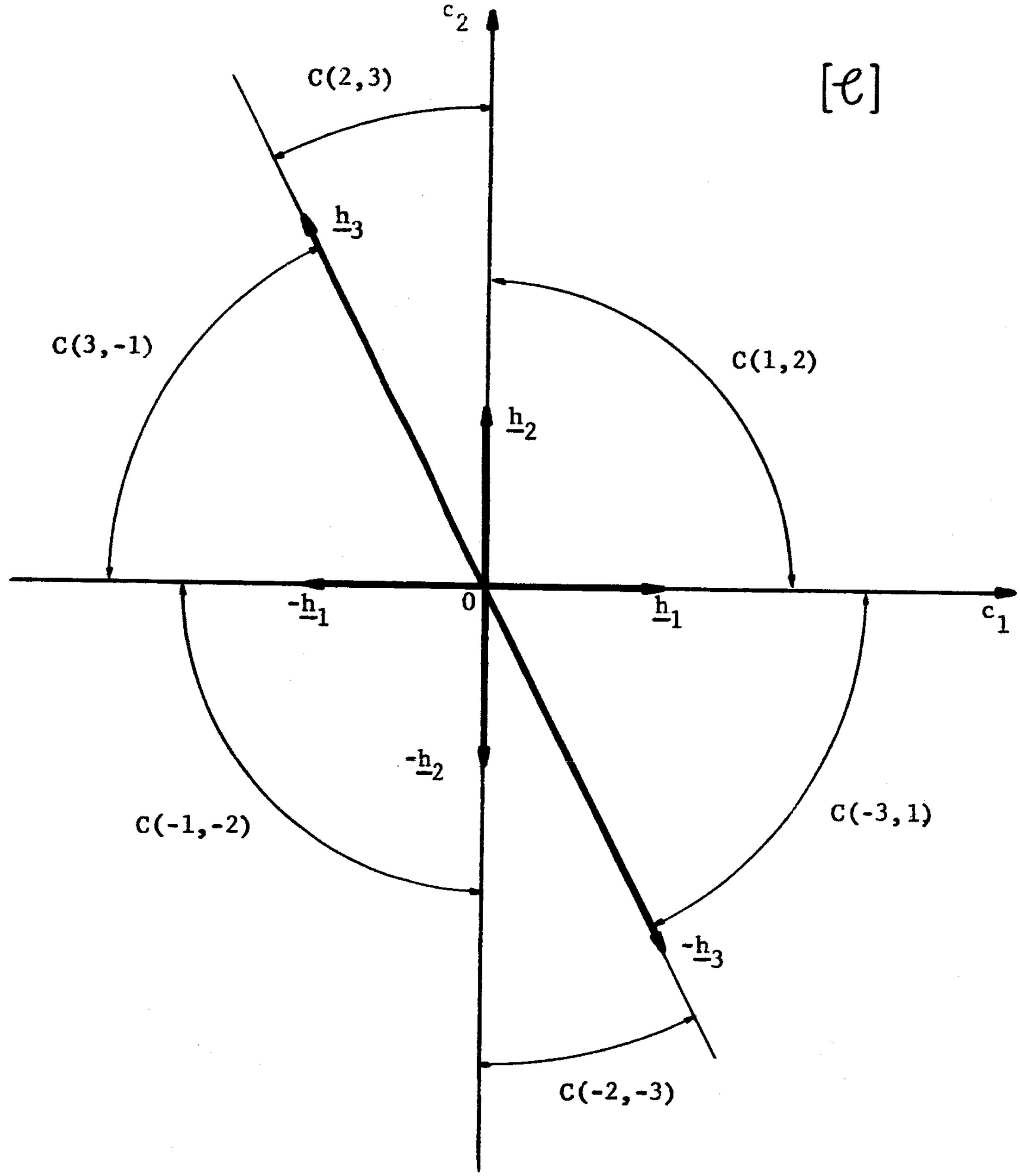


Figure 9. The cones $C(\pm i, \pm j)$, $i, j = 1, 2, 3$, for the plant $1/s^2$.

$$u(i) = \mu_1, \quad u(j) = \mu_2, \quad u(k) = 0, \quad k \neq i,j. \tag{2-66}$$

If the state lies in a cone where the sequence is not unique, this rule may not be the best choice because of practical considerations. The shape of the trajectory or the ease of synthesizing the input sequence will help to determine which sequence is to be chosen.

These results can be extended to n-th order systems with a regulation time of N sampling periods. The minimum fuel input sequence is obtained via consideration of the properties of the set $S_N(f)$.

General Properties of $S_N(f)$

$S_N(f)$ is defined as the set of all initial states that can be taken to the origin in N sampling periods with a fuel consumption $F \leqslant f$. Then

$$S_N(f) = \left(\underline{c} \,\middle|\, \underline{c} = \sum_{j=1}^{N} u(j)\, \underline{h}_j; \quad F = \sum_{j=1}^{N} \left|u(j)\right| \leqslant f \right) . \tag{2-67}$$

The following properties of $S_N(f)$ are proved by Lee and Desoer (22).

1. For f real and positive and for any integer N, $S_N(f)$ is a convex set, contains the origin as an interior point, and is symmetic with respect to the origin.

2. If $C_N(f)$ is the set of 2N points $f\underline{h}_j$, $-f\underline{h}_j$, j = 1, 2, ..., N, $S_N(f)$ is the convex hull (46, page 207) of $C_N(f)$. The convex hull of a set of points is, intuitively, the smallest convex set containing the points.

3. Let $\partial S_N(f)$ denote the boundary of $S_N(f)$. For a given $\underline{c}$, let f increase to f^* so that $\underline{c}$ is in $\partial S_N(f^*)$. Then f^* is the minimum fuel cost for the given $\underline{c}$.

4. If $\underline{c}$ is in $\partial S_N(f)$, $u(1)$, $u(2)$, ..., $u(N)$ is an optimal input sequence if and only if

$$\underline{c} = \sum_{j=1}^{N} u(j)\, \underline{h}_j \quad , \qquad \sum_{j=1}^{N} u(j) \;= f \quad . \tag{2-68}$$

5. Suppose $\underline{c}_1$, $\underline{c}_2$, ..., $\underline{c}_n$ are distinct points on $\partial S_N(f)$ and lie on a common supporting hyperplane, Φ . Let $\underline{c}$ be given by

$$\underline{c} = \sum_{k=1}^{n} \mu_k\, \underline{c}_k \quad ; \qquad \mu_k \geqslant 0 \ , \qquad \sum_{k=1}^{n} \mu_k = 1 \quad . \tag{2-69}$$

If $u_k(j)$, $j = 1, 2, \ldots, N$ are optimum for $\underline{c}_k$, $k = 1, 2, \ldots, n$, then

$$u(j) = \sum_{k=1}^{n} \mu_k\, u_k(j), \quad j = 1, 2, \ldots, N \quad , \tag{2-70}$$

are optimum for $\underline{c}$.

These properties can be used to generate an optimum input sequence. The method of synthesis in the general n-th order case will become evident after consideration of first and second order systems.

First Order Systems

The optimum input sequence for first order systems can be obtained without recourse to the properties of $S_N(f)$. For the first order plant

$$G_p(s) = \frac{1}{s + \lambda} \quad , \tag{2-71}$$

the invariant vectors are given by (Appendix A, Equation (A-49)),

$$\underline{h}_j = e^{(j-1)\lambda T} \quad , \quad j = 1, 2, \ldots, \tag{2-72}$$

which are scalars. The length of $\underline{h}_j$ is therefore

$$\ell_j = e^{(j-1)\lambda T} \quad , \quad j = 1, 2, \ldots \quad . \tag{2-73}$$

For stable plants $\lambda > 0$, and therefore, $\ell_j \geq \ell_k$ for $j > k$. If $\lambda = 0$, $\ell_1 = \ell_2 = \ldots = 1$. Figure 10 shows the invariant vectors for $\lambda > 0$.

The minimum fuel problem is to find what proportion, $u(j)$, of each invariant vector $\underline{h}_j$, $j = 1, 2, \ldots, N$, should be taken so that when they are added together, they reach $\underline{c}$ with the least cost F. In the linear case there is no limit on the amount of each that may be used. The solution is clearly to use only the longest available invariant vector to reach the initial state. For a given N, N = 1, 2, ..., if $\lambda > 0$ the longest vector is $\underline{h}_N$. If $\lambda = 0$, all the vectors are of unit length, and as long as all the $u(j)$ are of the same sign, it is immaterial how they are combined to reach $\underline{c}$. If $\lambda < 0$, corresponding to an unstable system, the first vector, $\underline{h}_1 = 1$ is the longest vector. The minimum fuel solutions are therefore:

1. $\lambda > 0$; the unique solution is

$$u(1) = u(2) = \ldots = u(N-1) = 0, \; u(N) = \underline{c}/e^{(N-1)\lambda T}. \tag{2-74}$$

2. $\lambda = 0$; there is no unique solution. Possible solutions are

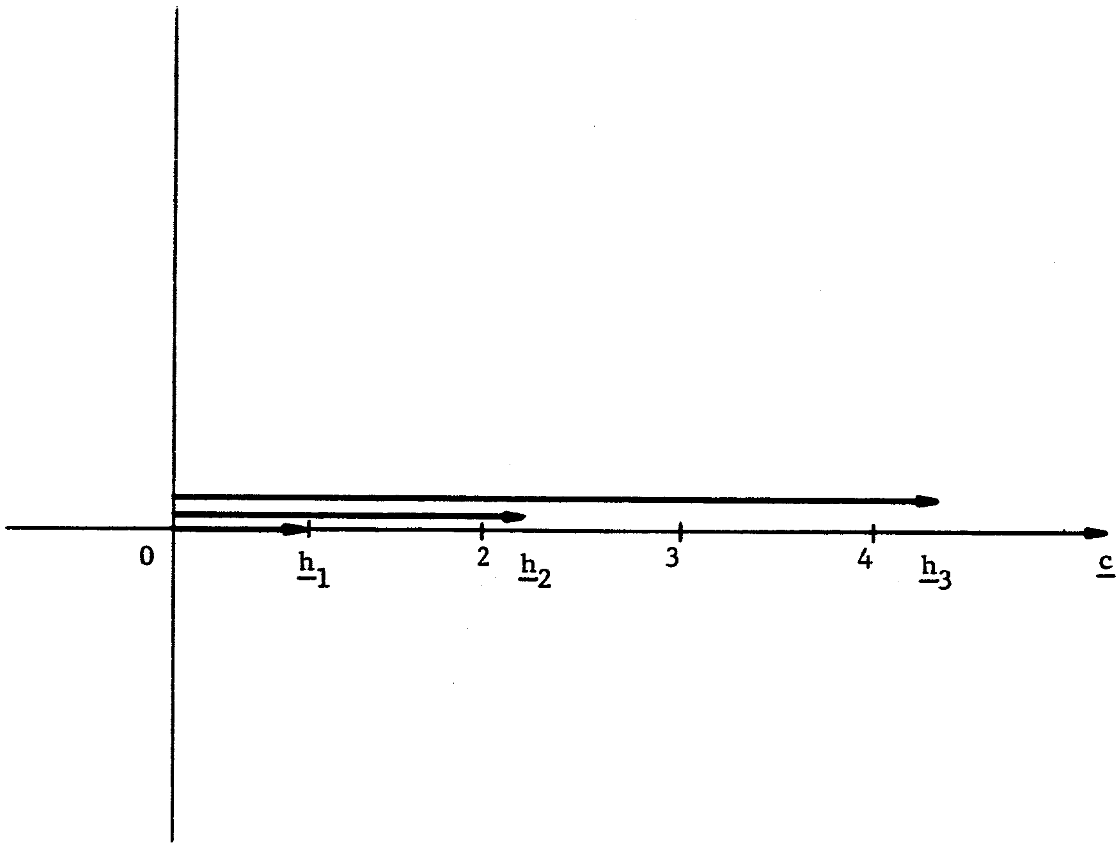


Figure 10. The invariant vectors for the first order system of Equation (2-71) with $\lambda > 0$, shown vertically displaced for clarity.

$$u(1) = u(2) = \dots = u(N) = \underline{c}/N \ , \qquad (2\text{-}75)$$

$$u(j) = \underline{c} \text{ for any one integer } j \text{ in } 1, 2, \dots, N \ . \qquad (2\text{-}76)$$

3. $\lambda < 0$; the unique solution is

$$u(1) = \underline{c}, \quad u(2) = u(3) = \dots = 0 \quad . \qquad (2\text{-}77)$$

These results make good sense when it is noted that with $\lambda > 0$, the state of the plant is moving into the origin of its own accord, and the longer it is allowed to do so, the less the cost of completing the regulation. With $\lambda < 0$, the free motion of the plant is away from the origin, so that the correcting force should be applied immediately. In this case the minimum fuel solution is also the minimum time solution. Except for the case $\lambda = 0$, pure integration, the solution is unique.

Second Order Systems

The set $S_N(1)$ is the convex hull of the set of 2N points $\underline{h}_j$, $-\underline{h}_j$, $j = 1, 2, \dots, N$ (see page 36). In general therefore, not all of the points $\pm \underline{h}_j$, $j = 1, 2, \dots, N$, will lie on $\partial S_N(1)$, the boundary of $S_N(1)$. It is necessary to distinguish between the invariant vectors that lie on $\partial S_N(1)$ and those that lie in the interior of $S_N(1)$. Therefore, let K denote the set of all distinct integers k such that $\underline{h}_k$ lies on $\partial S_N(1)$. Denote by p the number of members of K; there are therefore N - p invariant vectors in the interior of $S_N(1)$.

Let the 2p distinct line segments, $L_s(\pm i, \pm j)$, i, j in K, which together form the boundary of $S_N(1)$, generate the corresponding 2p cones $C_s(\pm i, \pm j)$. These cones cover the entire $\mathcal{C}$-space. Figure 11 shows these cones for a typical second order system with N = 5. The vectors $\underline{h}_1$ and $\underline{h}_2$ are shown interior to $S_5(1)$ and, therefore, p = 3.

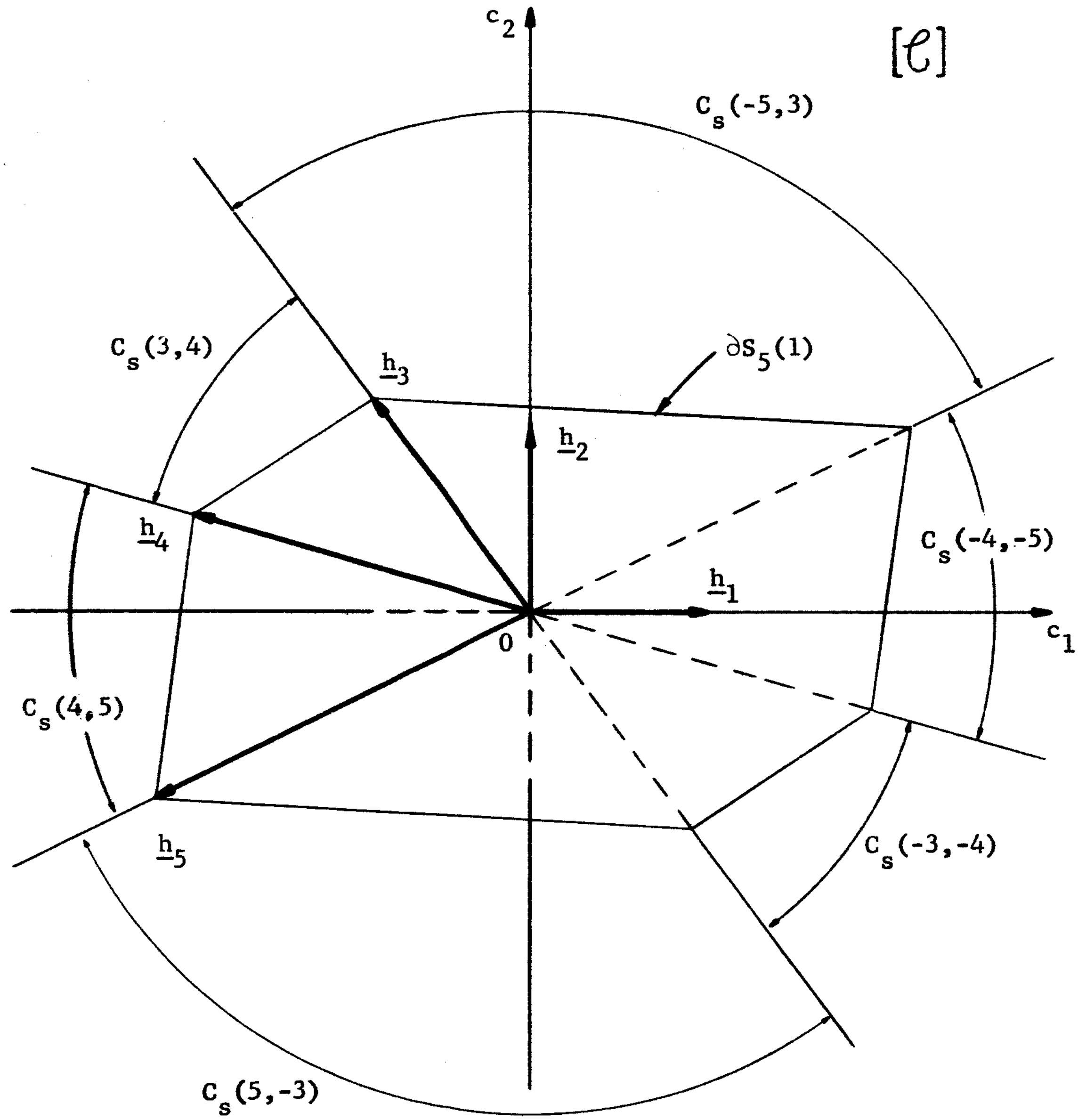


Figure 11. The cones $C_s(\pm i,\pm j)$ for a typical second order system.

Without loss of generality, it may be assumed, for notational convenience, that K contains the integers 1, 2, 3 and 4, and that $\underline{h}_1$, $\underline{h}_2$, $\underline{h}_3$ and $\underline{h}_4$ form adjacent cones $C_s(1, 2)$, $C_s(2, 3)$ and $C_s(3, 4)$ as shown in Figure 12. Furthermore, suppose that the initial state $\underline{c}$ lies in $\partial S_N(f)$ in the cone $C_s(2, 3)$. Then $\underline{c}$ can be represented as

$$\underline{c} = \mu_1 \underline{c}_1 + \mu_2 \underline{c}_2 \ , \quad \mu_1, \mu_2 \geqslant 0 \ , \quad \mu_1 + \mu_2 = 1 \ , \tag{2-78}$$

where

$$\underline{c}_1 = f\underline{h}_2 \ , \ \underline{c}_2 = f\underline{h}_3 \quad . \tag{2-79}$$

From Equation (2-68), optimum input sequences for the initial states $\underline{c}_1$ and $\underline{c}_2$ are seen to be, respectively,

$$u_1(j) = f\, \delta_{2j} \quad , \quad j = 1, 2, \ldots, N \ , \tag{2-80}$$

$$u_2(j) = f\, \delta_{3j} \quad , \quad j = 1, 2, \ldots, N \ , \tag{2-81}$$

where δ_{ij} is the Kronecker delta,

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad . \tag{2-82}$$

Therefore, from Equation (2-69),

$$u(1) = 0,\ u(2) = f\,\mu_1,\ u(3) = f\,\mu_2,\ u(4) = 0,\ \ldots,\ u(N) = 0 \tag{2-83}$$

is an optimum input sequence for the initial state $\underline{c}$ given by Equation (2-78).

More generally, if $\underline{c}$ is in the cone $C_s(\pm i, \pm j)$, i,j in K, and is represented by

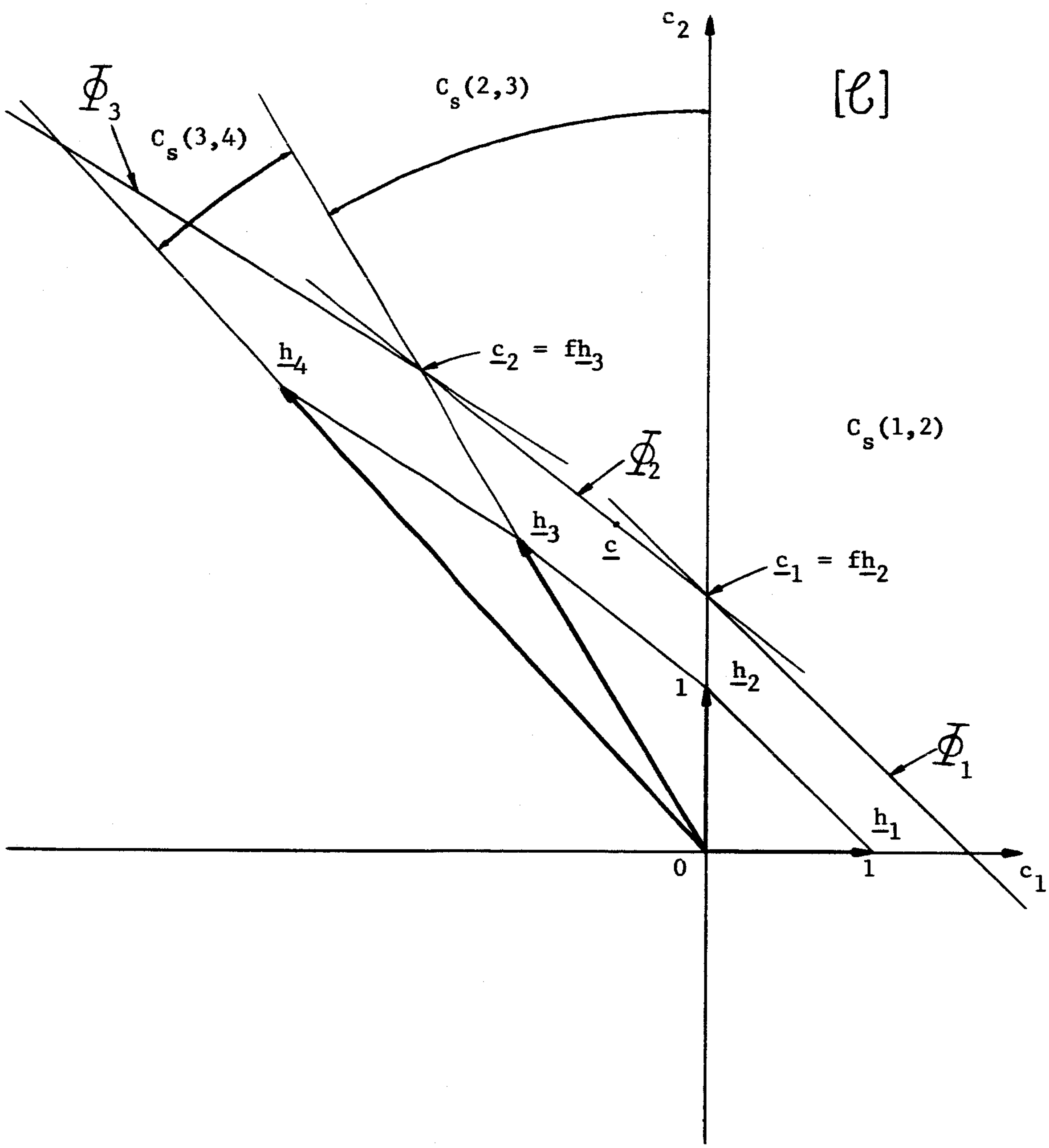


Figure 12. Examination of the fuel optimum sequence when the initial state $\underline{c}$ lies in the cone $C_s(2,3)$ on the line Φ_2.

$$\underline{c} = \mu_1 \underline{h}_i + \mu_2 \underline{h}_j \quad , \tag{2-84}$$

the minimum fuel input sequence is

$$u(i) = \mu_1 \ , \ \ u(j) = \mu_2 \ , \ \ u(k) = 0 \ , \ \ k \neq i,j \ . \tag{2-85}$$

The uniqueness of the input sequence is considered next. As shown in Figure 12, page 43, let Φ_1 be the line through $f\underline{h}_1$ and $f\underline{h}_2$, Φ_2 be the line through $f\underline{h}_2$ and $f\underline{h}_3$, and Φ_3 be the line through $f\underline{h}_3$ and $f\underline{h}_4$. Without loss of generality, as before let $\underline{c}$ lie in the cone $C_s(2, 3)$. Then the uniqueness of the optimum input sequence is given in the following theorem:

Theorem 1. For second order systems, with $\underline{c}$ in $C_s(2, 3)$, the minimum fuel input sequence is unique if, and only if,

$$\Phi_1 \neq \Phi_2 \neq \Phi_3 \ . \tag{2-86}$$

Proof. Suppose $\Phi_2 = \Phi_1$. Then $\underline{c}$ can be reached with minimum cost by using, in Equation (2-84), $\underline{h}_3$ and $\underline{h}_2$ or $\underline{h}_3$ and $\underline{h}_1$. Suppose $\Phi_2 = \Phi_3$. Here $\underline{c}$ can be reached with minimum cost by using either $\underline{h}_3$ and $\underline{h}_2$ or $\underline{h}_4$ and $\underline{h}_2$ in Equation (2-84). Therefore, necessity is proved.

Consider $\underline{c} = \underline{c}_1 = f\underline{h}_2$. Equation (2-80) gives an optimum input sequence $u(2) = f$, $u(j) = 0$, $j = 1, 3, \ldots, N$. This is the unique optimum control, since if $u(2)$ is less than f, and $\Phi_2 \neq \Phi_1$, $\underline{c}$ can only be reached by using additional invariant vectors, which gives a total fuel consumption greater than f. Similarly if $\underline{c} = \underline{c}_2 = f\underline{h}_3$, the optimum control sequence $u(3) = f$, $u(j) = 0$, $j = 1, 2, 4, \ldots, N$, is unique if $\Phi_2 \neq \Phi_3$. If $\underline{c}$ is given by Equation (2-78), the optimum input sequence of Equation (2-83) is unique if $\Phi_1 \neq \Phi_2 \neq \Phi_3$, since any other sequence gives a fuel cost greater than f (compare the introductory discussion on the minimum fuel problem). Thus sufficiency has been demonstrated.

Although the Theorem as stated is only concerned with uniqueness for second order systems, the extension of the Theorem to higher order systems is conceptually clear. In general $\partial S_N(f)$ is a polygon in n-dimensional $\mathcal{C}$ -space. Each face of the polygon has corners at the points $\pm f\underline{h}_k$, k in K. The initial state $\underline{c}$, in $\partial S_N(f)$, lies in one of these faces. Let this face be contained entirely in some hyperplane; i.e., any point in the face lies in the hyperplane, then the input sequence is unique if and only if no adjacent face is also contained entirely in the hyperplance.

The Theorem has immediate use. The synthesis of the control may be made easier by choosing one particular sequence from the alternative input sequences, and it may be that some additional performance criterion, such as time or energy, can be minimized to advantage. It is, therefore, of interest to know what plant pole arrangements give non-uniqueness.

Second order systems are now examined to ascertain when non-uniqueness can occur.

<u>Pole combinations for a non-unique input sequence</u>. If the poles of a second order system are real, the invariant vectors, $\underline{h}_j$, h = 3, 4, ..., lie in the second quadrant of $\mathcal{C}$ -space. With complex poles they can lie in any quadrant (Table I, Appendix A, page 252). Real poles will be discussed first, and then complex poles. Only stable systems will be considered; unstable systems can be treated in the same manner.

A. Real poles. Figure 13 shows the invariant vectors for a typical second order plant of the form

$$G_p(s) = \frac{1}{(s + \lambda_1)(s + \lambda_2)} \quad , \qquad \lambda_1, \lambda_2 > 0 \quad . \tag{2-87}$$

If the plant has one or two integrations, it is easily shown that the points $\underline{h}_j$, j = 1, 2, ..., lie on the line $c_1 + c_2 = 1$. Figure 14 shows the invariant vectors for the plant

$$G_p(s) = \frac{1}{s^2} \quad ; \tag{2-88}$$

Figure 15 shows them for the plant

$$G_p(s) = \frac{1}{s(s + \lambda_2)} \quad . \tag{2-89}$$

The cross-hatched regions in Figures 14 and 15 are, therefore, the regions for which the initial state has no unique optimum input.

If the plant does not have any integration; i.e., $\lambda_1, \lambda_2 > 0$, initial states with non-unique fuel optimum input sequences can still occur, although in a different manner. It can be shown that the slope of the line $\underline{h}_{j+1} - \underline{h}_j$, j = 1, 2, ..., becomes less negative as j increases: see the dashed line connecting $\underline{h}_1$, $\underline{h}_2$, $\underline{h}_3$, $\underline{h}_4$ and $\underline{h}_5$ in Figure 13. However, uniqueness depends on the shape of the boundary of $S_N(f)$, equivalently $S_N(1)$. Consider N = 3. While the point $\underline{h}_3$ cannot lie on the line $c_1 + c_2 = 1$, the point $-\underline{h}_3$ can, Figure 16 shows the arrangement and the corresponding set of initial conditions with non-unique input sequences. From Equation (A-50),

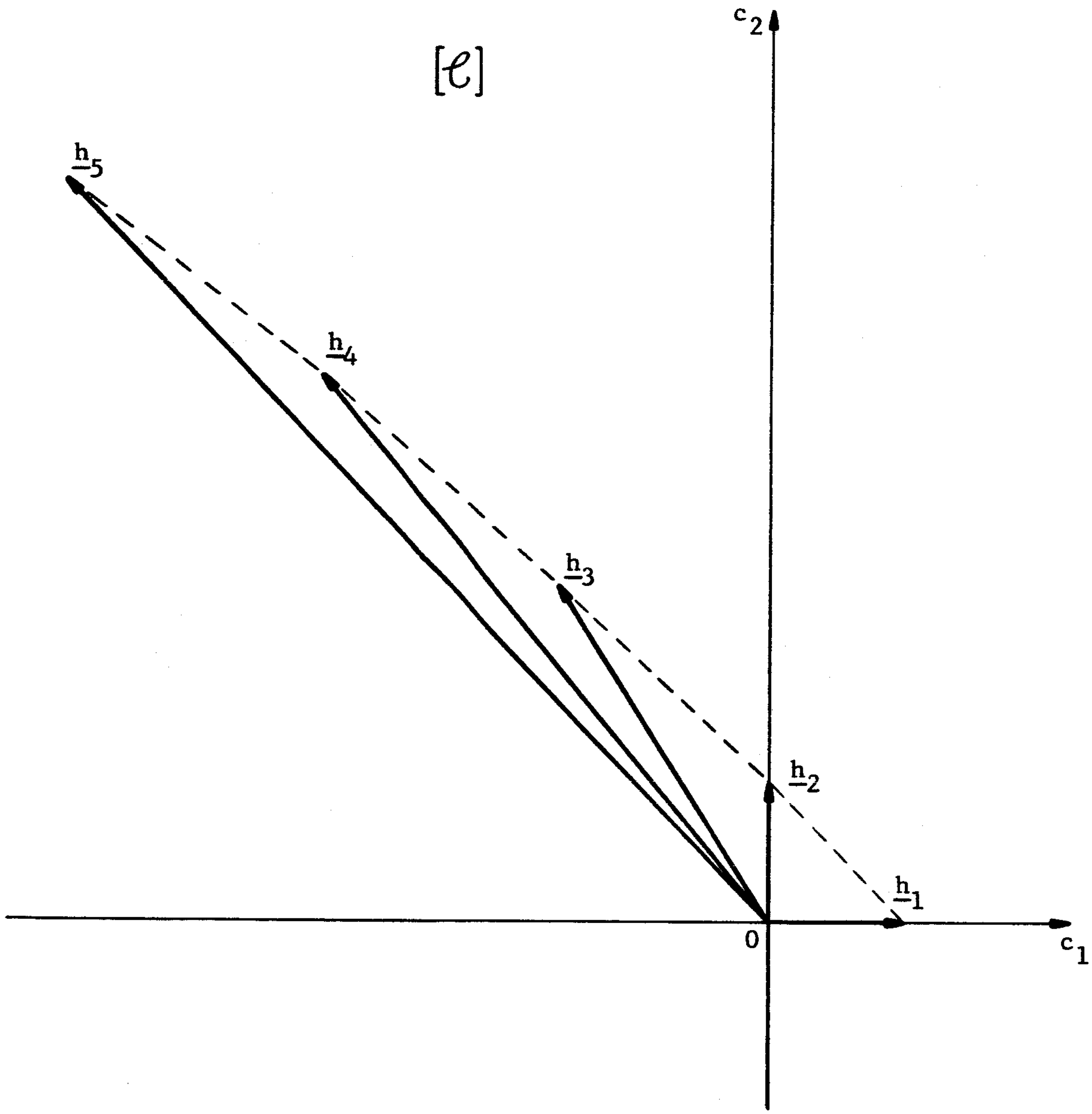


Figure 13. The invariant vectors for the plant of Equation (2-87).

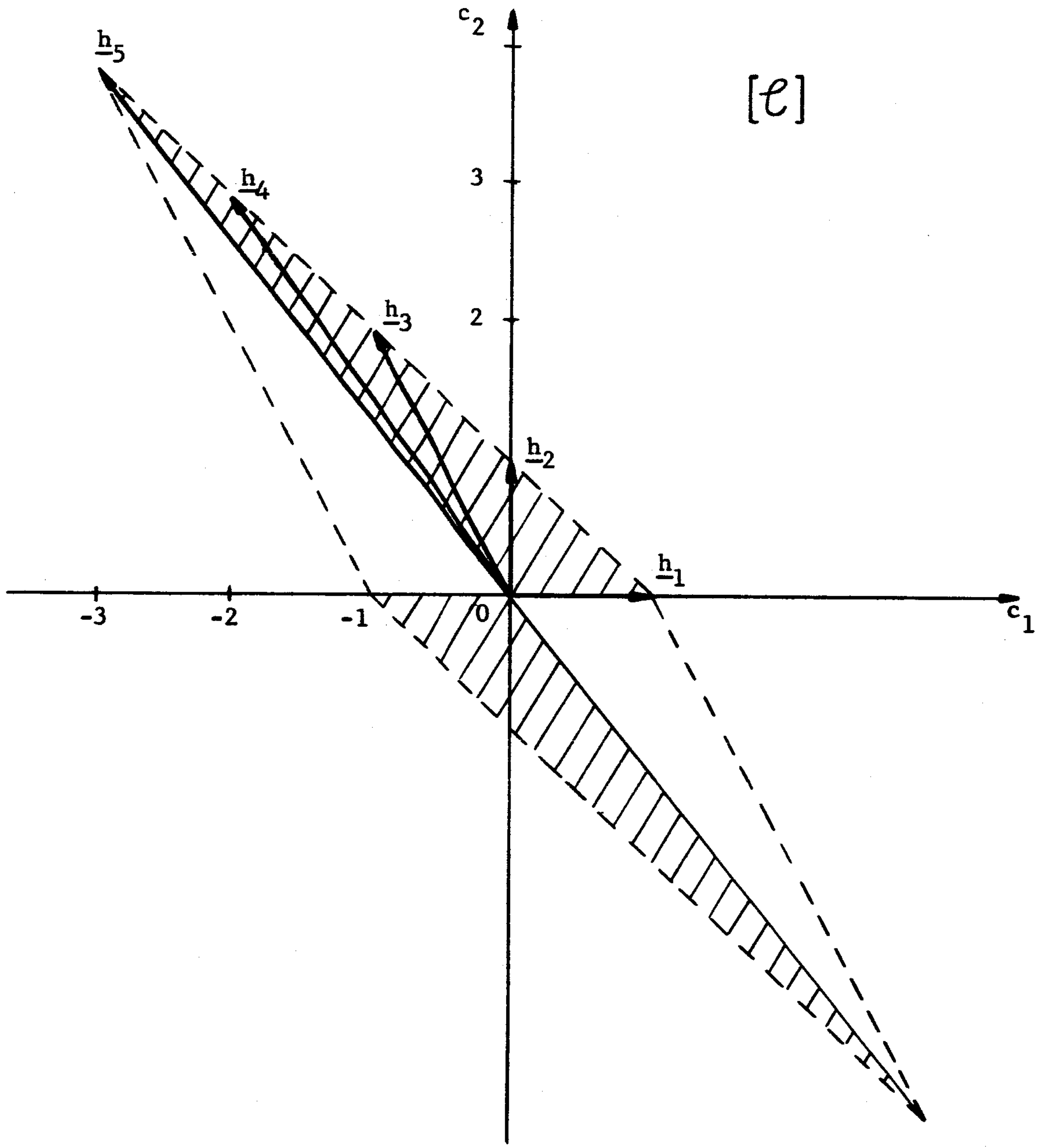


Figure 14. The invariant vectors for the plant of Equation (2-88).

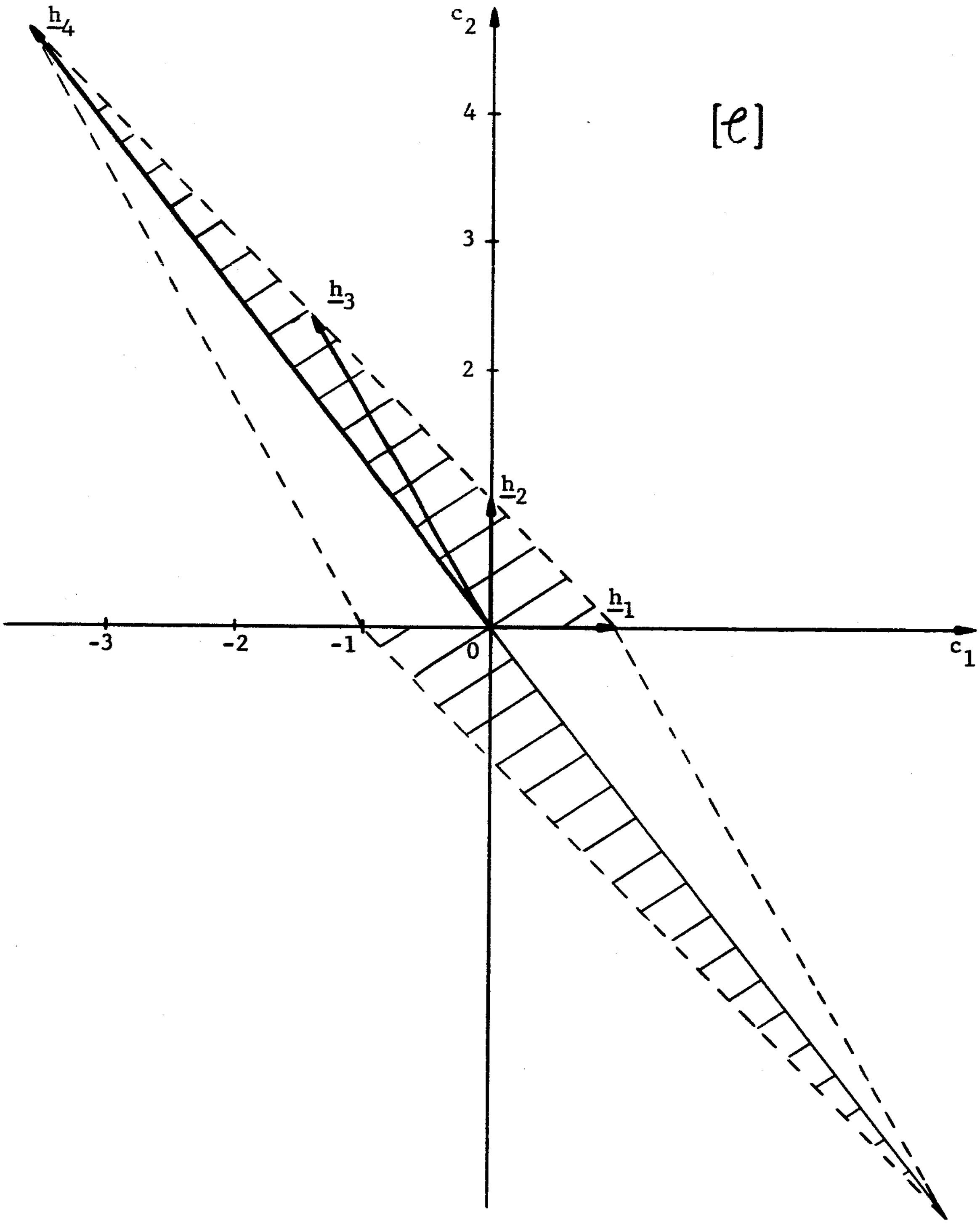


Figure 15. The invariant vectors for the plant of Equation (2-89).

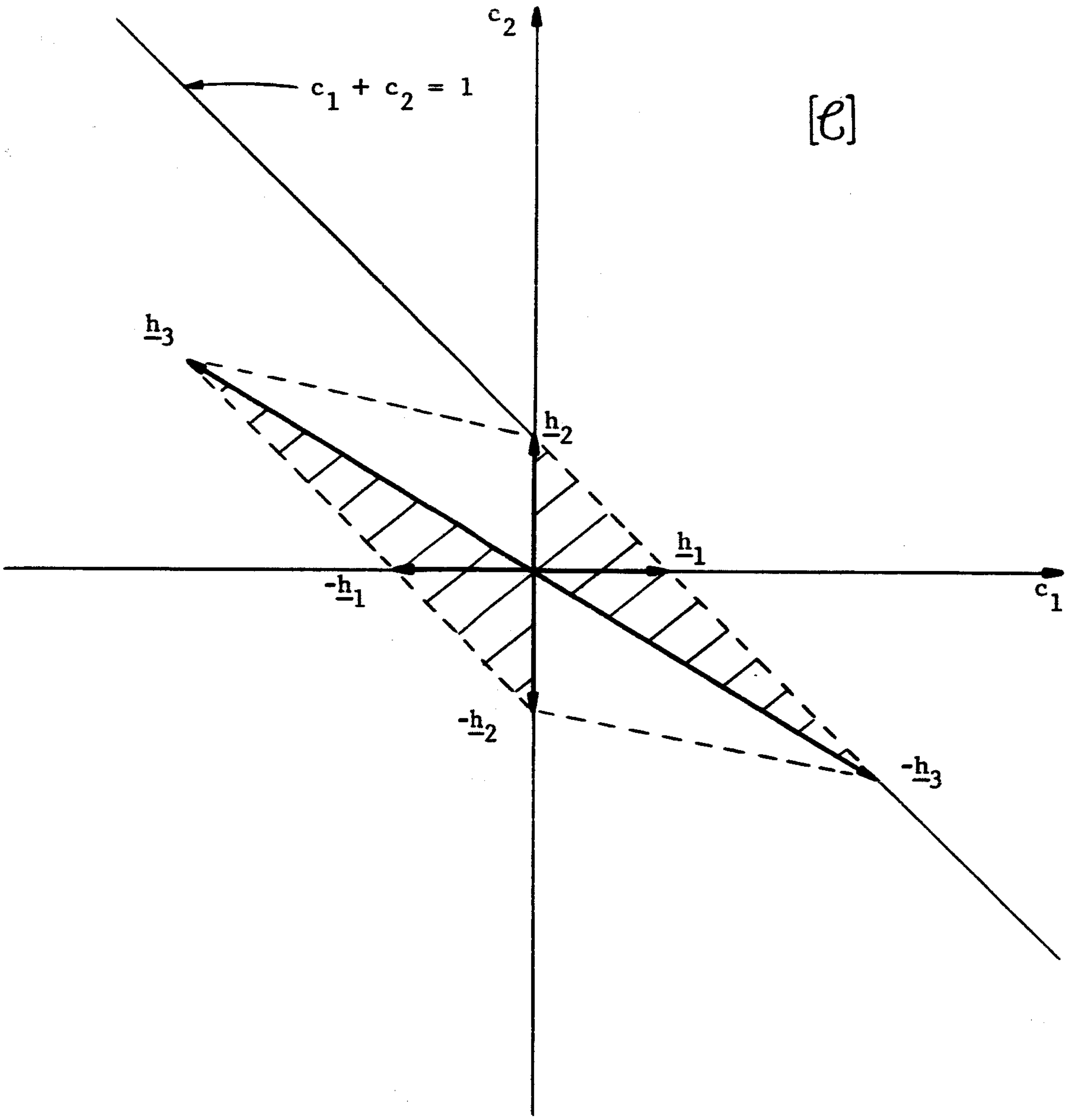


Figure 16. Initial states with a non-unique minimum fuel input sequence when $-\underline{h}_3$ lies on the line $c_1 + c_2 = 1$.

$$\underline{h}_3 = \begin{bmatrix} -e^{\gamma_1} e^{\gamma_2} \\ \\ e^{\gamma_1} + e^{\gamma_2} \end{bmatrix} \quad , \gamma_1 = \lambda_1 T, \ \ \gamma_2 = \lambda_2 T; \ \lambda_1, \lambda_2 > 0. \tag{2-90}$$

The condition for $-\underline{h}_3$ to lie on $c_1 + c_2 = 1$ is then

$$e^{\gamma_1} + e^{\gamma_2} - e^{\gamma_1} e^{\gamma_2} + 1 = 0 \quad . \tag{2-91}$$

Figure 17 shows the solution to Equation (2-91) in graphical form. The asymptotes $e^{\gamma_1} = 1$ and $e^{\gamma_2} = 1$ correspond to the cases where the plant has an integration.

If Equation (2-91) is satisfied by the plant and $N > 3$, the solution will always be unique since the $\underline{h}_j$, $j = 4, 5, \ldots$, for such a plant cannot lie on $c_1 + c_2 = 1$, and $\underline{h}_1$ will no longer lie on $\partial S_N(1)$. Figure 18 illustrates this for $N = 4$.

In general, for a given N, non-uniqueness for second order plants given by Equation (2-87) occurs if $-\underline{h}_N$ lies on the line $c_1 + c_2 = 1$. The condition for $-\underline{h}_N$ to lie on $c_1 + c_2 = 1$, $N = 3, 4, \ldots$, is obtained with the help of Equations (A-52) and (A-53). The plant poles must be such that, for $N = 3, 4, \ldots$,

$$1 + \sum_{i=0}^{N-2} e^{(N-2-i)\gamma_2 + i\gamma_1} - \sum_{i=0}^{N-3} e^{(N-2-i)\gamma_2 + (i+1)\gamma_1} = 0. \tag{2-92}$$

When $\gamma_1 = \gamma_2 = \gamma$, Equation (2-92) reduces to

$$(N - 2)e^{(N-1)\gamma} - (N - 1)^{(N-2)\gamma} - 1 = 0, \quad N = 3, 4, \ldots \quad . \tag{2-93}$$

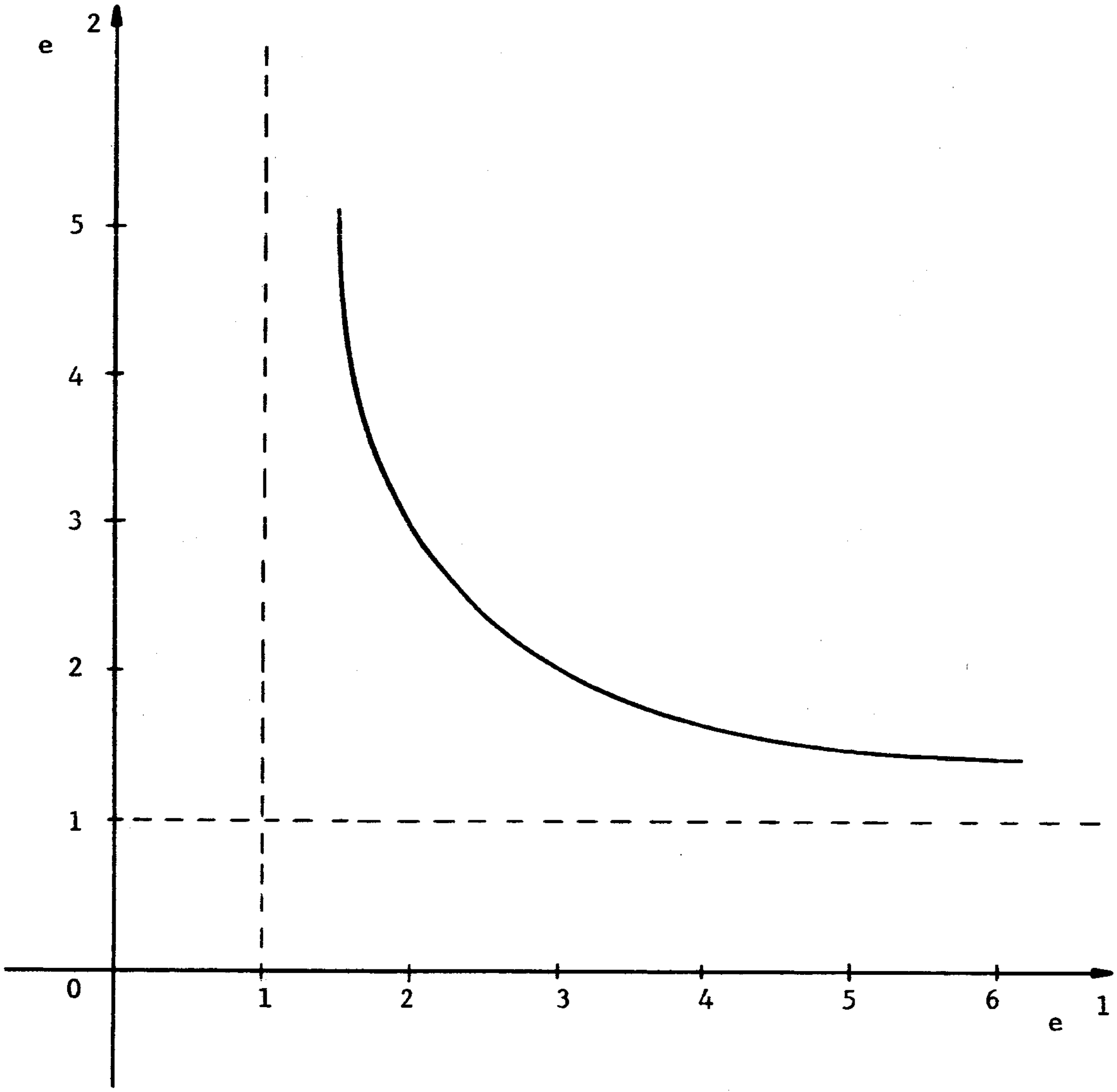


Figure 17. Graphical solution to the non-uniqueness of Equation (2-91).

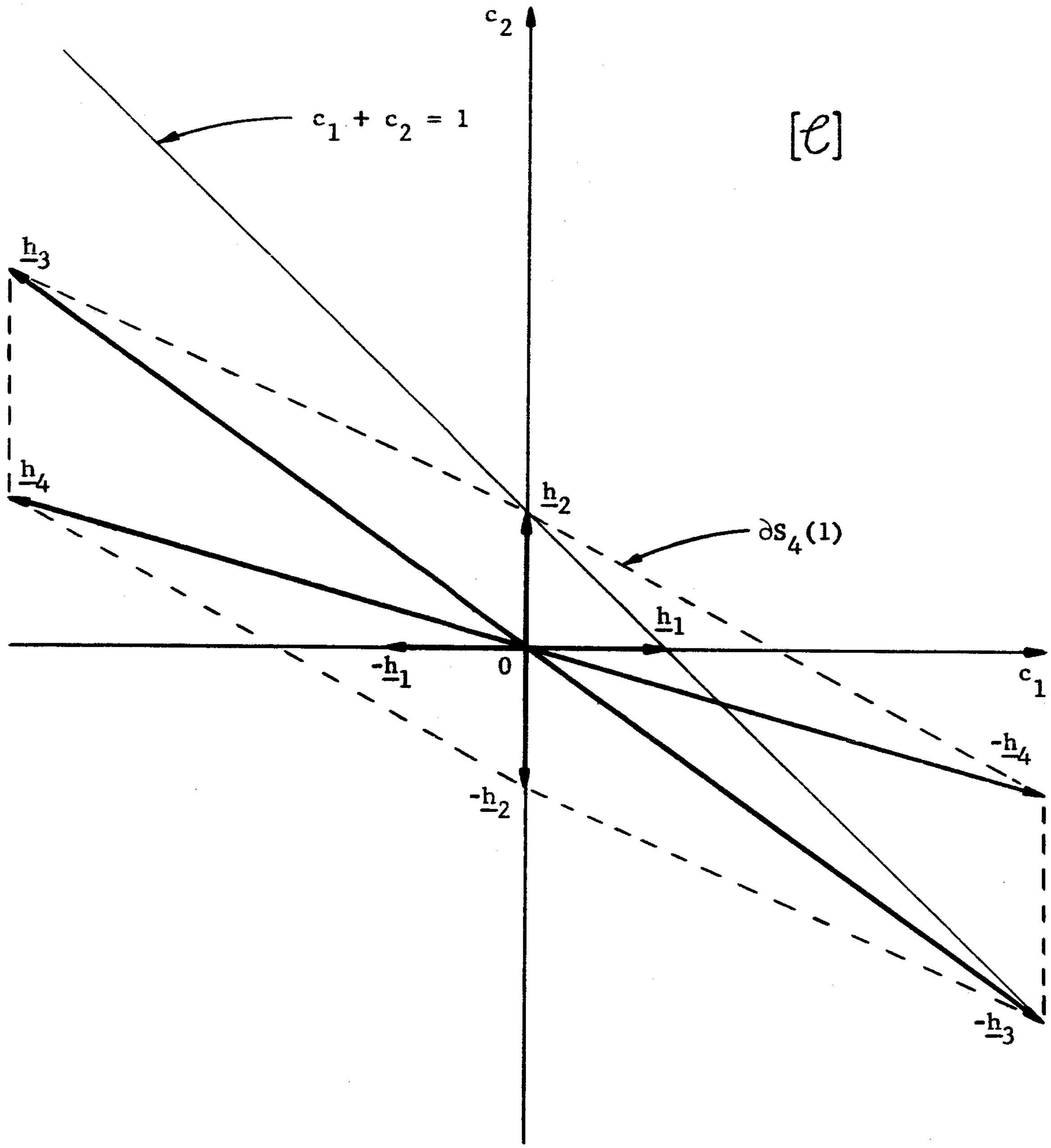


Figure 18. An example of the case when only three of the invariant vectors lie on the line $c_1 + c_2 = 1$.

When $N = 3$, the value of $\gamma > 0$ which satisfies Equation (2-93) is $\gamma = \log_e(1 + \sqrt{2})$. As N increases the corresponding γ decreases in a strictly monotonic manner.

B. <u>Complex poles</u>. With α and β real, let

$$e^{\gamma_1} = \alpha + j\beta, \quad e^{\gamma_2} = \alpha - j\beta \quad . \tag{2-94}$$

From Equation (2-91), the condition for $\pm \underline{h}_3$ to lie on $c_1 + c_2 = 1$ is then

$$\alpha^2 + \beta^2 - 2\alpha \pm 1 = 0 \quad . \tag{2-95}$$

Therefore, $+ \underline{h}_3$ lies on $c_1 + c_2 = 1$ when

$$(\alpha - 1)^2 + \beta^2 = 0 \quad , \tag{2-96}$$

and $- \underline{h}_3$ lies on $c_1 + c_2 = 1$ when

$$(\alpha - 1)^2 + \beta^2 = 2 \quad . \tag{2-97}$$

Since α and β are real, Equation (2-96) has no solution. The non-uniqueness, therefore, occurs when Equation (2-97) is satisfied by the plant poles. The semi-circle of Figure 19 shows the permissible values of α and β that do satisfy Equation (2-97).

If $N > 3$, the problem of finding pole locations which give non-uniqueness is complicated by the fact that the invariant vectors can lie in any of the four quadrants of $\mathcal{C}$-space. However, second order plants can always be checked for uniqueness by actually drawing the set $S_N(1)$ for the given plant poles.

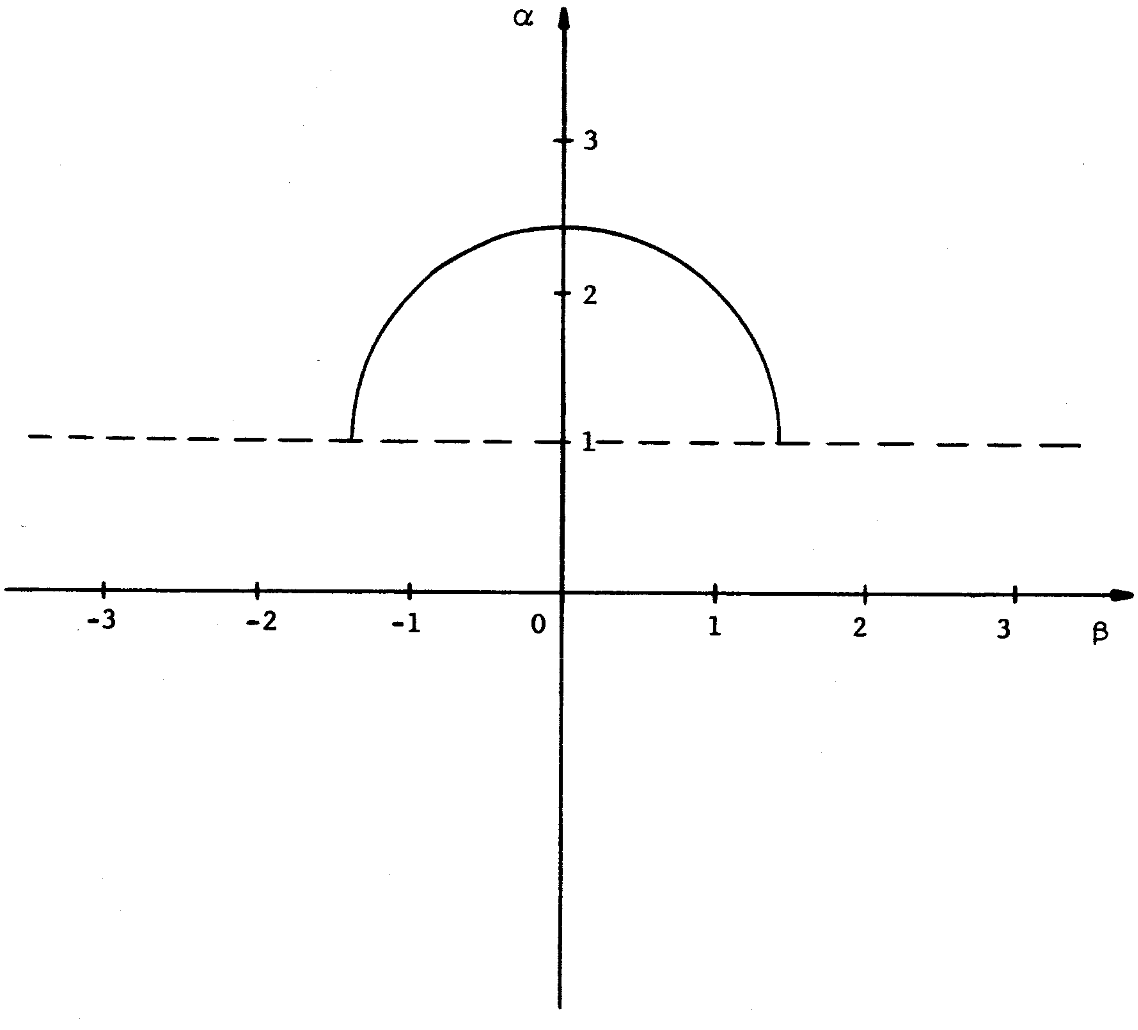


Figure 19. The values of α and β that give non-uniqueness of the minimum fuel input sequence when the plant has complex poles, for the case N = 3.

V. EXAMPLES

Two simple examples will now be discussed in order to show how the minimum fuel and energy theory may be applied in practice.

Consider a trolley of unit mass rolling on rails, propelled by either a battery driven d.c. motor or by gas jets. At time $t = 0$, the position and velocity of the trolley are given as $x_1(0)$ and $x_2(0)$ respectively. The trolley is to be brought to rest in no more than four seconds; i.e., $x_1(4) = x_2(4) = 0$.

The driving force (torque) at the driven wheels is directly proportional to the current supplied by the battery; the energy supplied by the battery is proportional to the square of the current. With the jet propelled system, ejecting gas at a fixed nozzle velocity, the driving force is proportional to the rate at which mass is ejected. In the case of the battery driven trolley, the least energy is to be used in bringing the trolley to rest. With the jet-propelled trolley, the gas consumption is to be minimized.

The differential equation describing the motion of the trolley is

$$\frac{d^2x_1}{dt^2} = u(t) \quad , \tag{2-98}$$

where, in the battery powered case $u(t)$, the driving force, is proportional to the current, and in the jet powered case, $u(t)$ is proportional the the rate of mass (gas) ejected.

The energy supplied by the battery is proportional to

$$\int_0^4 u^2(t)\, dt \quad , \tag{2-99}$$

and the mass of gas expelled is proportional to

$$\int_0^4 |u(t)|\, dt \quad . \tag{2-100}$$

The control, $u(t)$, is required to be the output of a sample-hold device, with a sampling period of T seconds. The vector difference equation describing, at each sampling period, the motion of the trolley under the influence of piecewise constant inputs is therefore, (see Appendix A)

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix} u(k) \; . \tag{2-101}$$

The position and velocity of the trolley at the k-th sampling instant are respectively $x_1(k)$ and $x_2(k)$. The canonical vectors, defined by Equation (A-19), are

$$\underline{r}_j = \begin{bmatrix} (2j-1)\,\frac{T^2}{2} \\ -T \end{bmatrix} \quad , \qquad j = 1, 2, \ldots, \tag{2-102}$$

so that, from Equation (2-8),

$$R = \begin{bmatrix} \frac{T^2}{2} & \frac{3}{2}\,T^2 \\ -T & -T \end{bmatrix} \quad . \tag{2-103}$$

Let the sampling period be T = 1, and let the given initial state be

$$x(0) = \begin{bmatrix} 2 \\ -2 \end{bmatrix} . \qquad (2\text{-}104)$$

Figure 20 shows the relationship between the canonical vectors and $\underline{x}(0)$ in the state space. Only the first four canonical vectors are needed, since with T = 1, N = 4. The dashed line in Figure 20 shows how the trolley would move if no control forces were applied. Equation (2-14) transforms the initial state $\underline{x}(0)$ into $\mathcal{C}$ -space, giving

$$\underline{c} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} . \qquad (2\text{-}105)$$

Figure 21 shows $\underline{c}$ and the first four invariant vectors in $\mathcal{C}$ -space.

Minimum Energy Example

The optimum input sequence can be obtained by using either the graphical technique described earlier in this chapter, or the minimum energy equations, Equations (2-27) and (2-25). The graphical method is used first.

Figure 22 shows, in $\mathcal{A}$-space, the lines $\underline{h}_j^t\ \underline{a}^o = 1$, $\underline{h}_j^t\ \underline{a}^o = 0$ for j = 1, 2, 3 and 4. The line $u^o(1) = 1$ in $\mathcal{C}$ -space is found by transforming graphically the points A and B from $\mathcal{A}$ into $\mathcal{C}$. Other points on $\underline{h}_1^t\ \underline{a}^o = 1$ could be used; however, A and B are perhaps the most convenient. Figure 23 shows, by means of the dashed construction lines, how the points A' and B' are generated from A and B. For example, B' is found by adding $b/(a + b)\ \underline{h}_3$ to the point B (a and b are shown in

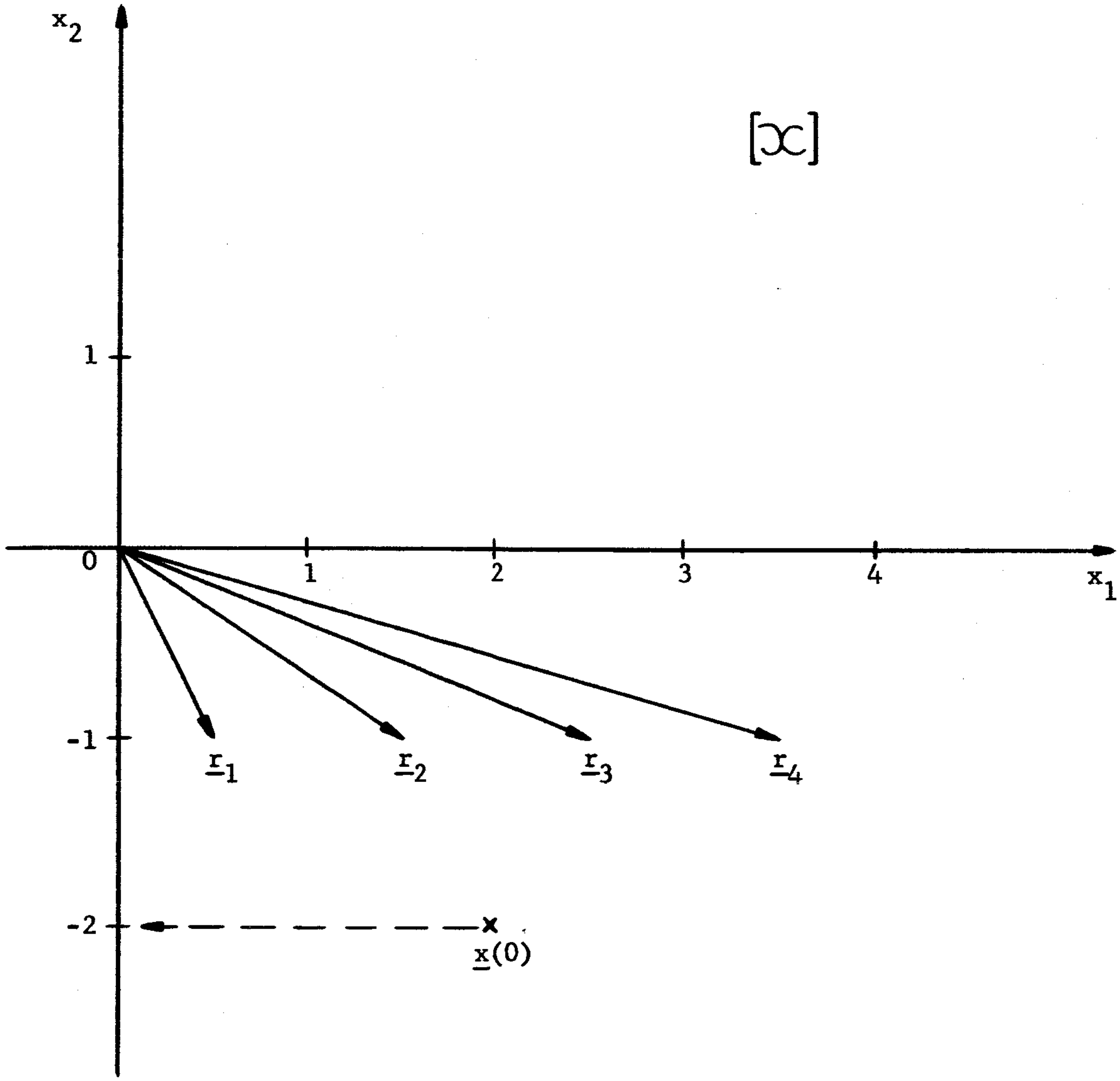


Figure 20. The canonical vectors and the initial state $\underline{x}(0)$ for the plant of Equation (2-98).

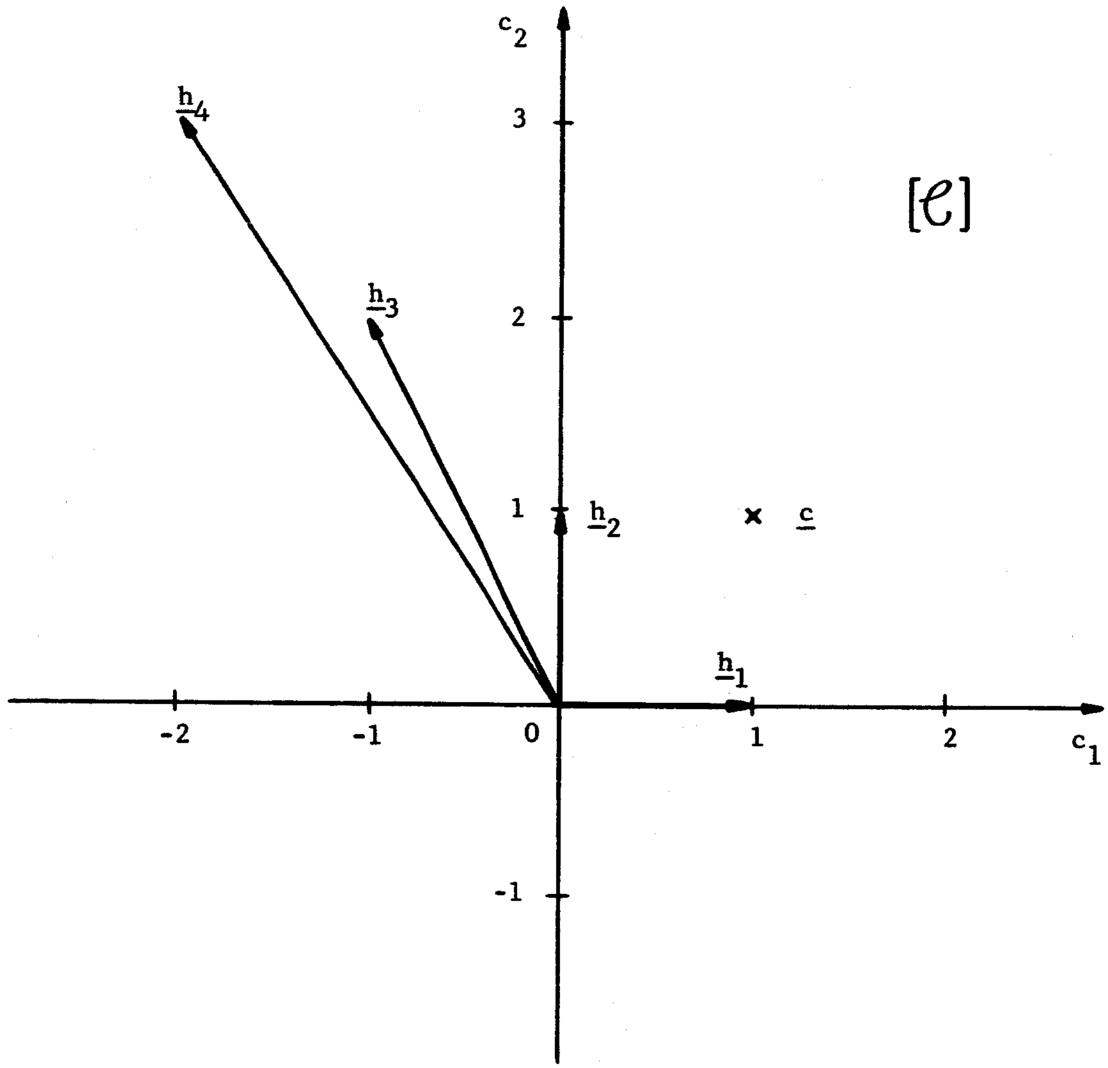


Figure 21. The invariant vectors and the initial state $\underline{c}$ for the plant of Equation (2-98).

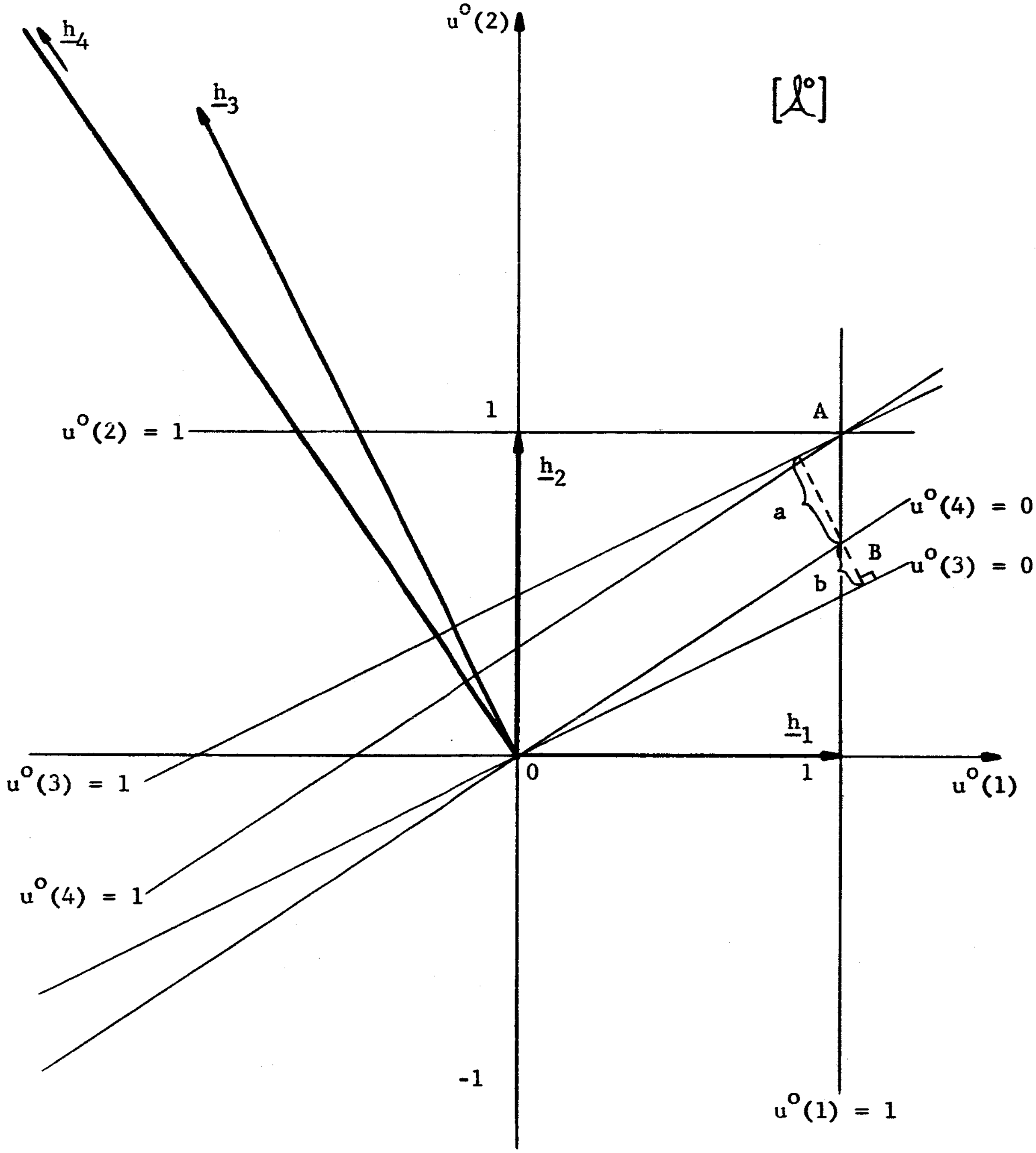


Figure 22. The lines $u^o(j) = 1$, $j = 1, 2, 3, 4$ and $u^o(j) = 0$, $j = 3, 4$ in $\mathcal{X}^o$ -space for the plant of Equation (2-98).

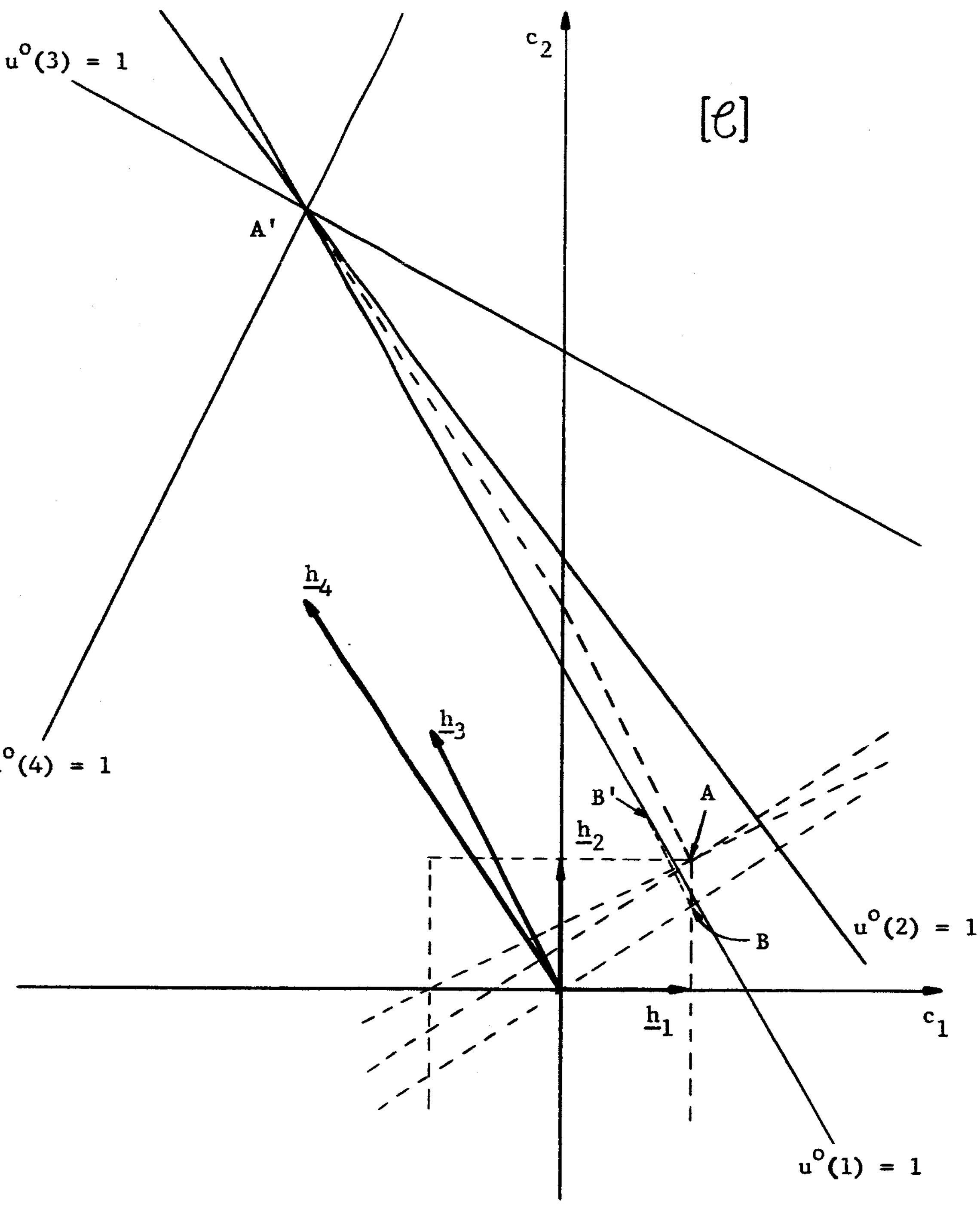


Figure 23. The generation of the lines $u^o(j) = 1$ in $\mathscr{C}$-space for the plant of Equation (2-98).

Figure 22, page 61). Figure 22, in fact, is not actually needed for the generation of the lines $u^o(j) = 1$ in $\mathcal{C}$-space: as mentioned earlier, the coordinates of $\mathcal{C}$-space can serve the dual purpose of supporting both $\mathcal{A}$-space and $\mathcal{C}$-space.

The lines $u^o(2) = 1$, $u^o(3) = 1$ and $u^o(4) = 1$ are obtained in a similar fashion and are also shown in Figure 23. Therefore, by interpolation and extrapolation, the optimum input sequence can be estimated. The approximate input sequence is, therefore,

$$u^o(1) \simeq 1.1 \ , \quad u^o(2) \simeq 0.68 \ , \quad u^o(3) \simeq 0.3 \ , \quad u^o(4) \simeq -0.1 \ . \tag{2-106}$$

Equations (2-27) and (2-25) will now be used to calculate the exact optimum input sequence. From Equation (2-17), the matrix H is

$$H = \begin{bmatrix} -1 & -2 \\ 2 & 3 \end{bmatrix} , \tag{2-107}$$

and

$$\left[I + HH^t\right]^{-1} = \begin{bmatrix} 0.7 & 0.4 \\ 0.4 & 0.3 \end{bmatrix} . \tag{2-108}$$

Therefore, from Equation (2-27),

$$\underline{a}^o = \begin{bmatrix} u^o(1) \\ u^o(2) \end{bmatrix} = \begin{bmatrix} 0.7 & 0.4 \\ 0.4 & 0.3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.1 \\ 0.7 \end{bmatrix} \tag{2-109}$$

and from Equation (2-25),

$$\underline{b}^o = \begin{bmatrix} u^o(3) \\ \\ u^o(4) \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1.1 \\ \\ 0.7 \end{bmatrix} = \begin{bmatrix} 0.3 \\ \\ -0.1 \end{bmatrix} . \qquad (2\text{-}110)$$

The graphical method compares quite well with the exact calculation. The energy cost, from Equation (2-39), is

$$E^o = \underline{c}^t \, \underline{a}^o = 1.8 \quad . \qquad (2\text{-}111)$$

Minimum Fuel Example

Figure 24 shows the set $S_4(1)$. The minimum fuel input sequence is not unique since $\underline{c}$ lies in the cone $C_s(1,2)$. The time optimum minimum fuel input sequence is clearly obtained when $\underline{c}$ is represented by $\underline{h}_1$ and $\underline{h}_2$:

$$u(1) = 1, \quad u(2) = 1, \quad u(3) = 0, \quad u(4) = 0 \; . \qquad (2\text{-}112)$$

The other two possible input sequences are obtained when $\underline{c}$ is represented by either $\underline{h}_1$ and $\underline{h}_3$, or by $\underline{h}_1$ and $\underline{h}_4$. The sequences are respectively;

$$u(1) = 1.5, \quad u(2) = 0, \quad u(3) = 0.5, \quad u(4) = 0, \qquad (2\text{-}113)$$

$$u(1) = 5/3, \quad u(2) = 0, \quad u(3) = 0, \quad u(4) = 1/3. \qquad (2\text{-}114)$$

The fuel cost is F = 2.

In conclusion, Figure 25 shows the four trajectories in the state space $\mathcal{X}$. Trajectory (a) is the minimum energy trajectory, and trajectories (b), (c) and (d) are the three possible minimum fuel trajectories.

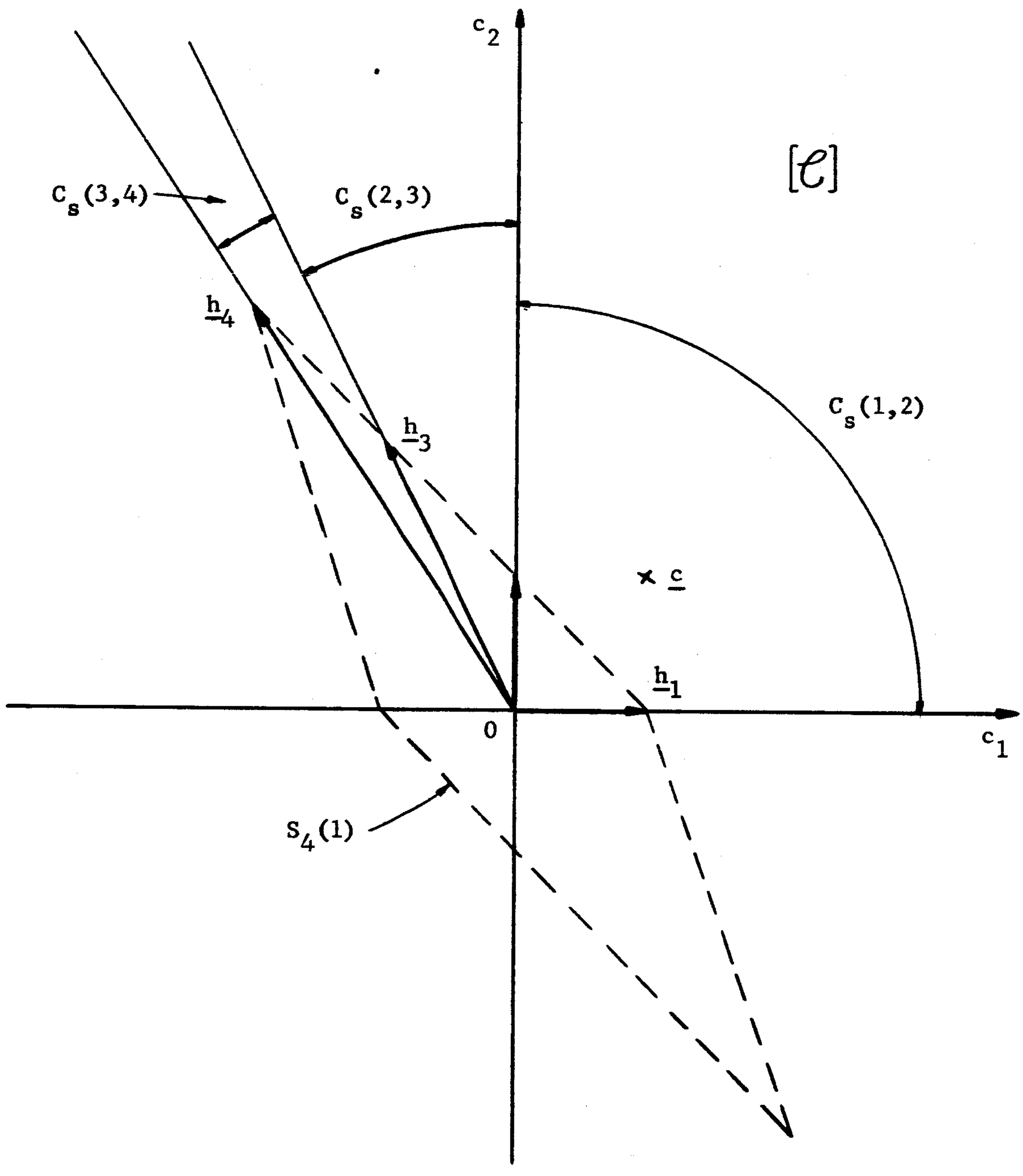
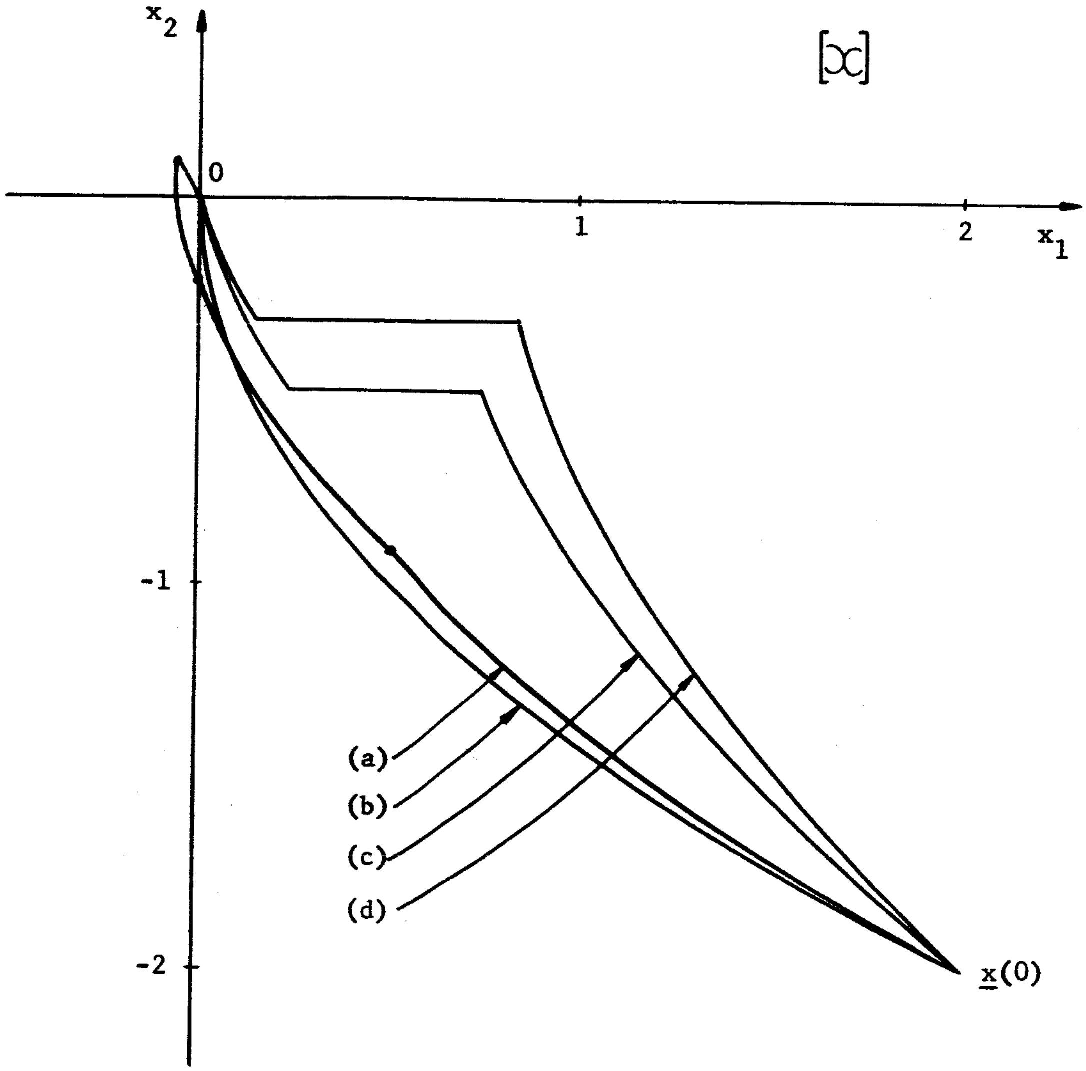


Figure 24. The set $S_4(1)$ for the plant of Equation (2-98).



(a) Minimum energy trajectory, Equations (2-98) and (2-99)
(b) Minimum fuel trajectory from Equation (2-101)
(c) Minimum fuel trajectory from Equation (2-102)
(d) Minimum fuel trajectory from Equation (2-103)

Figure 25. Minimum energy and minimum fuel trajectories for the trolley example.

CHAPTER III

THE MINIMUM ENERGY PROBLEM WITH INPUT SATURATION

I. INTRODUCTION

The minimum energy problem with input saturation is, in general, very complex. In order to introduce some of the problems associated with amplitude constrained inputs without confusing the issue with complex notation, first order systems are discussed initially. The discussion is largely intuitive, and leads to an algorithm for generating the optimum constrained input sequence in an open loop manner.

In general, if the initial states lies sufficiently close to the origin of $\mathcal{C}$ -space, the problem is solved. However, there is a substantital region of initial states for which one or more members of the corresponding linear minimum energy input sequence exceeds the saturation limits. By working in a partitioned correction space, rather than the solution space, several properties of the optimum constrained input sequence can be derived. It is shown that the minimum energy problem with amplitude constrained control is equivalent to finding which members of the input sequence are to be set equal to the saturation limit. Theorems 2 and 3 are helpful in finding such members.

While for first order systems the problem is solved, second order systems are solved in general only if the plant has integration or if it has complex poles. General results have not been obtained for other plants.

II. FIRST ORDER SYSTEMS

In Chapter II, the linear minimum fuel problem was approached by first considering the properties of the optimum input sequence for first order plants. With first order systems, the invariant vectors lie on the real line, and it is for this reason that intuition may be employed to advantage. First order plants, while of interest in themselves, can again be used to throw some light on the general problem of optimum regulation with saturation.

Suppose, for the moment, that the saturation constraint is relaxed. For a first order system, with a pole at $s = -\lambda$, the invariant vectors are scalars:

$$\underline{h}_{j+1} = e^{j\gamma}, \quad j = 0, 1, \ldots, \tag{3-1}$$

where $\gamma = \lambda T$. Using Equation (3-1),

$$\left[I + HH^t\right]^{-1} = 1 \Big/ \sum_{j=0}^{N-1} e^{2j\gamma}, \tag{3-2}$$

where, with n = 1, the n x N - n matrix H is defined in Equation (2-17). The optimum input sequence, in the absence of the saturation constraint, is then given by Equations (2-27) and (2-25):

$$u^o(j+1) = e^{j\gamma}\, \underline{c} \Big/ \sum_{k=0}^{N-1} e^{2k\gamma}, \quad j = 0, 1, \ldots, N-1, \tag{3-3}$$

where $\underline{c}$, again a scalar quantity, is the initial state in $\mathcal{C}$-space, corresponding to the state $\underline{x}(0)$ in $\mathcal{X}$-space, and is given by Equation (2-14). Consider the properties of this input sequence: from Equation (3-3),

$$|u^o(1)| \leqslant |u^o(2)| < \dots \leqslant |u^o(N)| \quad ; \quad \gamma > 0 \tag{3-4}$$

$$|u^o(1)| > |u^o(2)| \quad \dots > |u^o(N)| \quad ; \quad \gamma < 0 \ . \tag{3-5}$$

It is interesting to note that, for $\underline{c} \neq 0$, none of the members of the minimum energy input sequence can equal zero, while in the minimum fuel problem all but one of the input members were equal to zero (page 39). The linear minimum energy and fuel problems are alike in that, for $\gamma > 0$, the last member, $u^o(N)$, is the largest, and for $\gamma < 0$ the first member, $u^o(1)$, is the largest. For a given N, Figure 26 shows how $u(j)$ increases linearly with $\underline{c}$ for the three typical cases, $\gamma > 0$, $\gamma = 0$ and $\gamma < 0$. There is no loss of generality in confining the discussion to stable systems; i.e., $\gamma \geqslant 0$. From Equation (3-3), if

$$0 \leqslant \underline{c} \leqslant e^{-(N-1)\gamma} \sum_{j=0}^{N-1} e^{2j\gamma} \quad , \tag{3-6}$$

then $0 \leqslant u^o(N) \leqslant 1$. Therefore, for a given N, if $\underline{c}$ satisfies Equation (3-6), the optimum input sequence satisfies the saturation constraint. It is convenient to define the set M_N: for first order systems M_N is the set of all states $\pm\, \underline{c}$, where $\underline{c}$ satisfies Equation (3-6). Thus, if $\underline{c}$ does not lie in M_N, the optimum input sequence given by Equation (3-3) has one or more members which exceed the saturation limits. In this

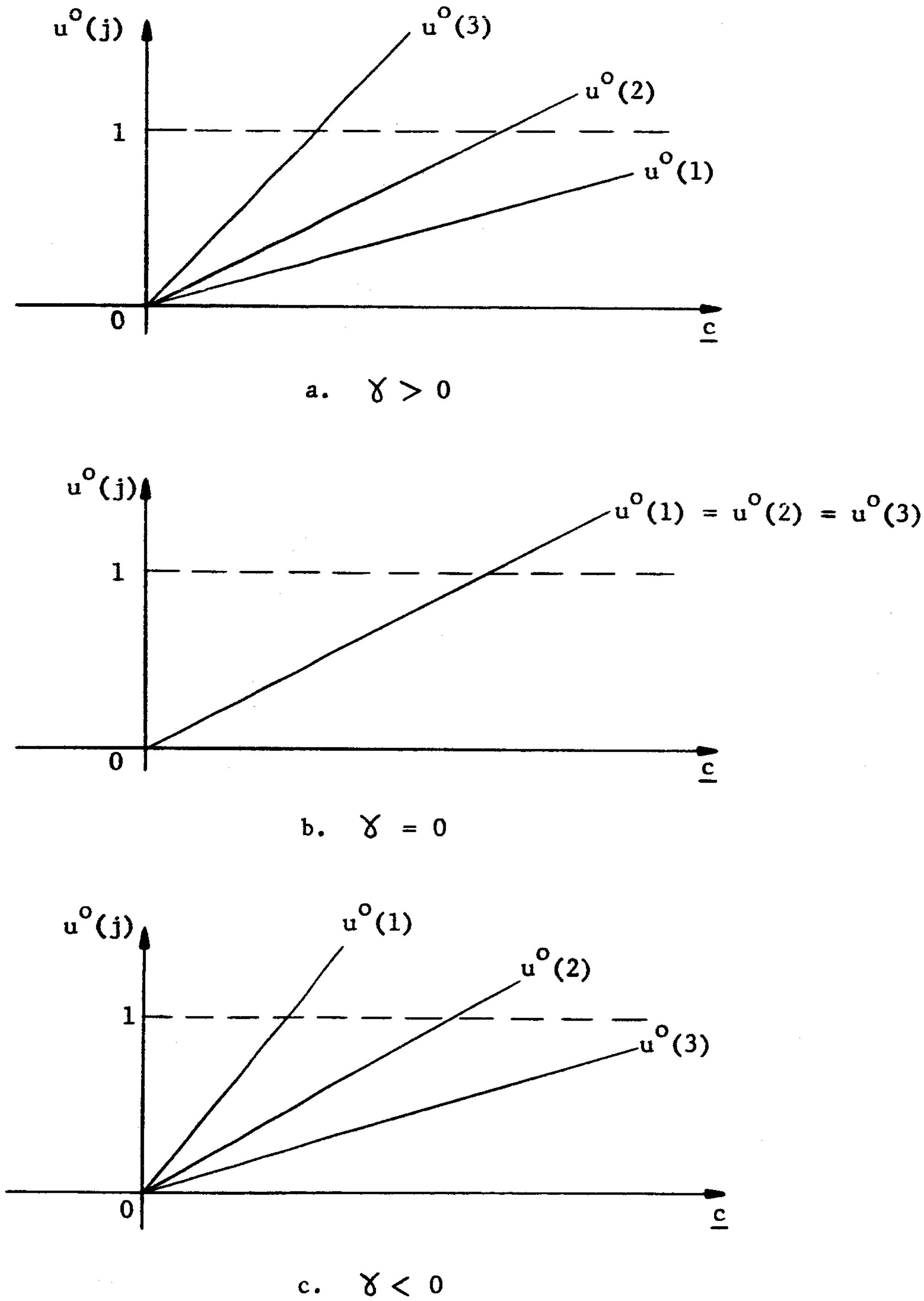


Figure 26. The inputs $u^o(j)$ as a function of the initial state for three representative first order systems.

case, the set Γ_N provides the answer to the question: is there an optimum input sequence which does satisfy the saturation constraint? Such a sequence does indeed exist, if and only if the initial state lies in the set Γ_N. The statement that $\underline{c}$ is in Γ_N is a compact way of saying that $\underline{c}$ can be represented by

$$\underline{c} = \sum_{j=1}^{N} u(j)\, \underline{h}(j) \ , \qquad |u(j)| \leqslant 1,\ j = 1, 2, \ldots, N, \tag{3-7}$$

so that there is a solution, u(1), ..., u(N), to the deadbeat regulator problem. For first order systems, the set Γ_N is the set of all initial states $\pm\ \underline{c}$, where

$$0 \leqslant \underline{c} \leqslant \sum_{j=0}^{N-1} e^{j\gamma} \ . \tag{3-8}$$

For $e^{\gamma} = 2$, Figure 27 shows M_3 and Γ_3. Figure 27 also shows $u^o(1)$, $u^o(2)$ and $u^o(3)$ as a function of $\underline{c}$. The following analogy is helpful in appreciating why $\underline{c}$ must satisfy Equation (3-8) in order for there to be a solution to the deadbeat regulator problem.

The vector $u(j)\ \underline{h}_j$ can be imagined to be a telescoping rod, which may be extended from zero length ($u(j) = 0$) up to a maximum length, the length of $\underline{h}_j$ ($u(j) = \pm 1$). The deadbeat regulator problem, the problem of representing $\underline{c}$ in the form of Equation (3-7), can be considered as follows: for a given N, N rods are available, each of which, for $\gamma \neq 0$, has a different maximum length. The end of one rod being fixed at the origin of $\mathcal{C}$-space, the rods are to be placed end to end so

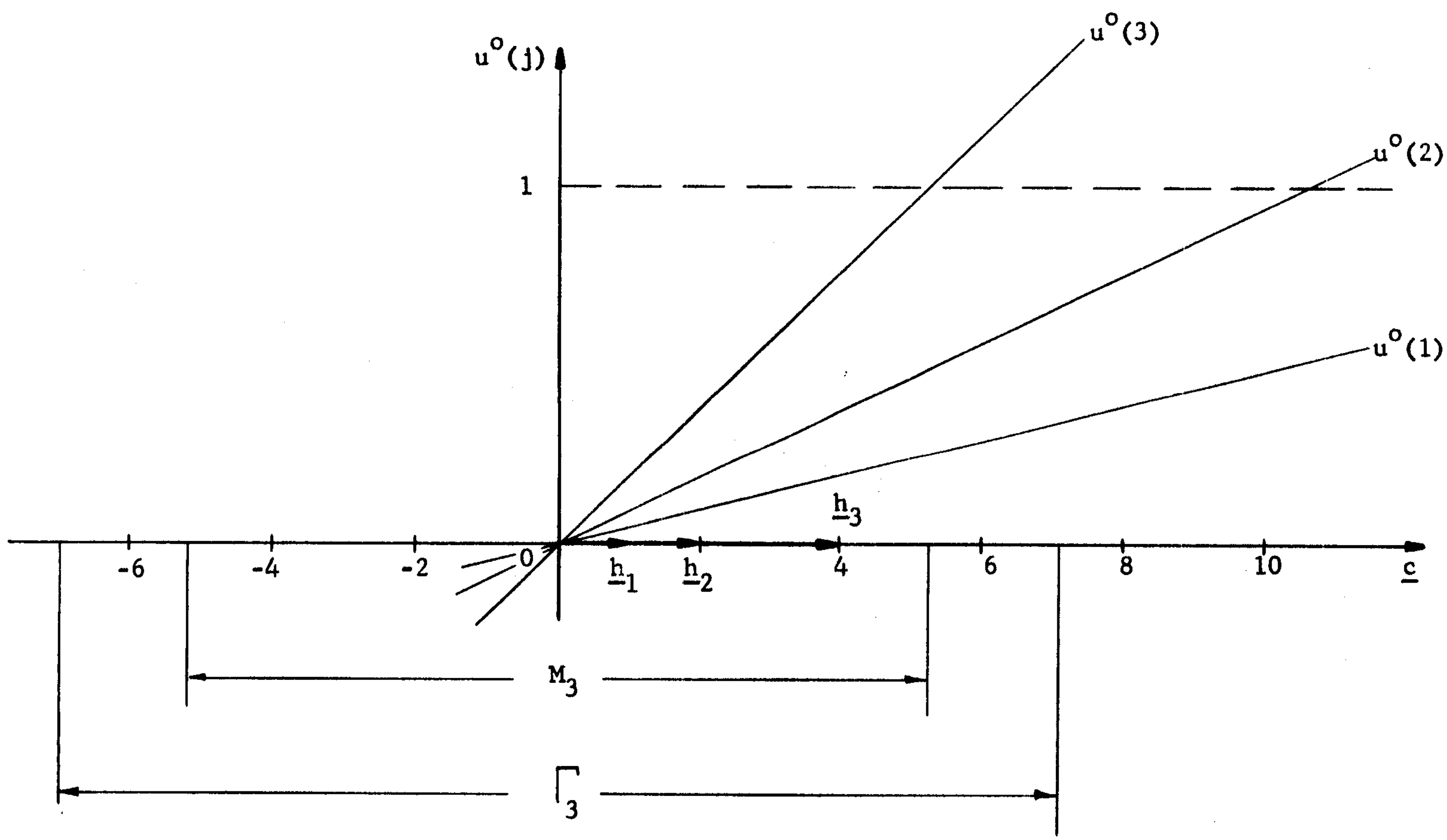


Figure 27. The sets M_3 and Γ_3 for a first order system with $e^{\gamma} = 2$.

that their resultant combination reaches the given state $\underline{c}$. If, having used all the available rods at their maximum extensions, it is still not possible to reach $\underline{c}$, then there is no solution to the problem. In order to bring $\underline{c}$ to the origin, either N must be increased (more rods must be made available) or the saturation constraint must be made less stringent (the maximum allowable length of the rods must be increased). The set Γ_N is simply the largest set of states that can be reached by combining together the N rods; for example, Figure 82 in Appendix A, page 244, illustrates how Γ_3 is formed in this way for a second order system.

Assuming $\underline{c}$ is not in M_N, but that N is large enough so that $\underline{c}$ is in Γ_N, there remains the problem of finding what amplitude constrained input sequence minimizes the energy, E. If $\underline{c}$ is not in M_N, it is of interest to know how many of the members of the input sequence exceed the saturation limit. From Equation (3-3), with $\gamma > 0$, if

$$e^{-j\gamma} \sum_{k=0}^{N-1} e^{2k\gamma} < |\underline{c}| \leqslant e^{-(j-1)\gamma} \sum_{k=0}^{N-1} e^{2k\gamma} , \qquad (3\text{-}9)$$

then $u^o(1)$, ..., $u^o(j)$ do not exceed the saturation limits and $u^o(j+1)$, ..., $u^o(N)$ do exceed the saturation limits.

Let the j-th member of the minimum energy amplitude constrained input sequence be $u^e(j)$. It is postulated now, and verified later, page 121, that, having calculated $u^o(j)$ from Equation (3-3),

$$\text{if} \quad |u^o(j)| > 1 \quad \text{then} \quad u^e(j) = \text{sgn. } u^o(j) , \qquad (3\text{-}10)$$

where

$$\text{sgn. } u^c(j) = \begin{cases} 1 & \text{if } u^o(j) > 0 \\ -1 & \text{if } u^o(j) < 0 \end{cases} . \qquad (3\text{-}11)$$

In words, Equation (3-10) states, "If the j-th member of the unconstrained optimum sequence exceeds the saturation constraint, the corresponding member of the optimum constrained input sequence is set equal to the saturation limit." The telescoping rod analogy can be used to show that the postulate is intuitively reasonable. If $\underline{c}$ is not in M_N, Equation (3-4) shows that at least $|u^o(N)| > 1$. This is to be expected since $\underline{h}_N$ is the longest invariant vector; under the minimum energy criterion (as under the minimum fuel criterion) $\underline{h}_N$ would, therefore, be utilized the most in reaching $\underline{c}$. Furthermore, when the inputs are constrained in amplitude, $|u(j)| \leqslant 1$, it seems reasonable to expect that $u^e(N)$ should be reduced as little as possible from $u^o(N)$; i.e., $u^e(N) = 1$.

The solution space (31) gives an alternate method of showing that the postulate is correct, at least for N = 2. Figure 28 shows the two-dimensional solution space. Input sequences that satisfy the saturation constraint are represented by points on or within the square centered on the origin. Figure 28 also shows two circles centered on the origin. Points on the circles represent input sequences with equal energy cost: the larger the circle, the larger the cost. Input sequences that take an initial state $\underline{c}$ into the origin of $\mathcal{C}$-space must satisfy the deadbeat constraint, Equation (A-59), which, for N = 2 and n = 1 gives

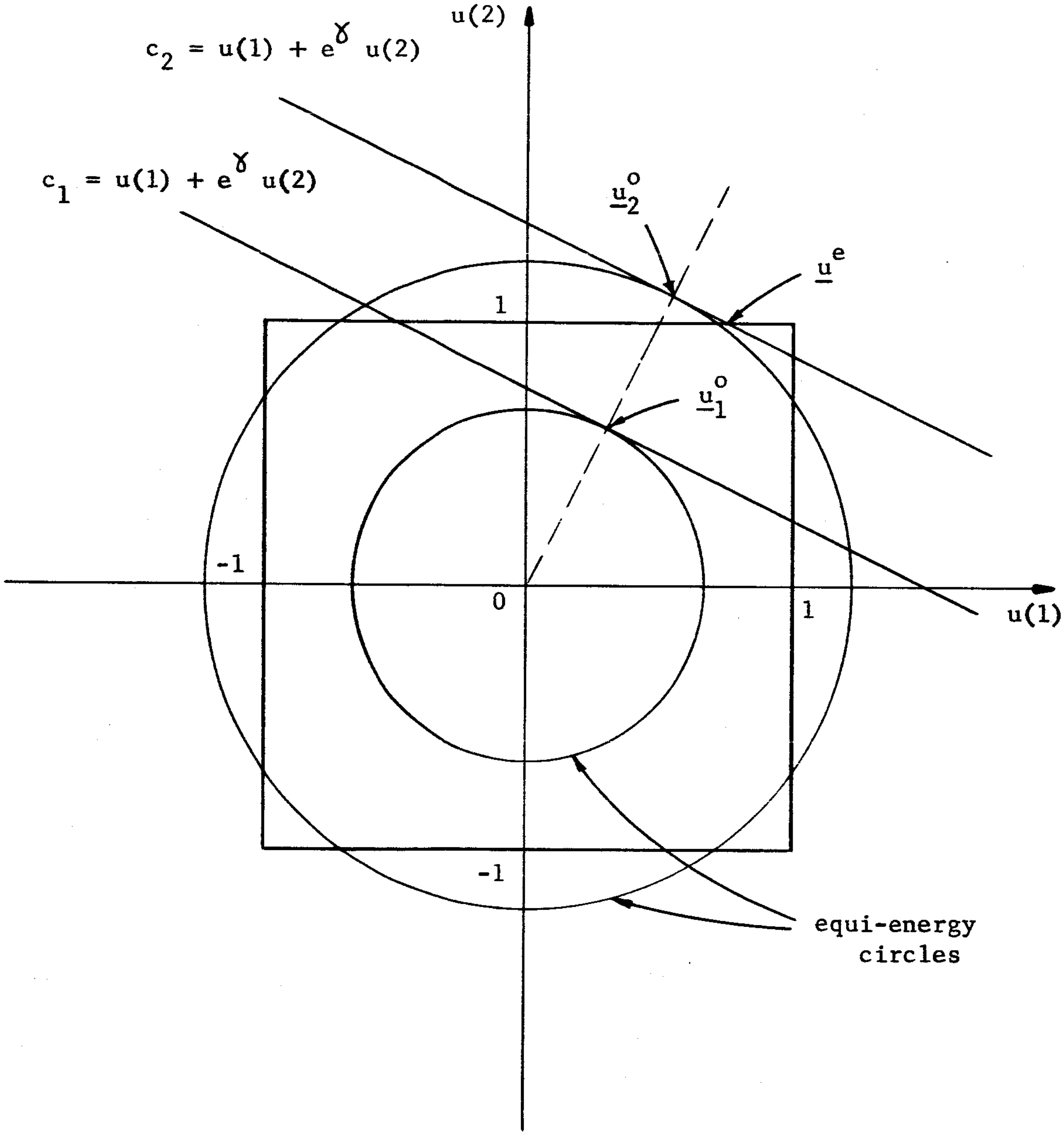


Figure 28. The minimum energy problem for a first order system in a two-dimensional solution space.

$$\underline{c} = u(1)\, \underline{h}_1 + u(2)\, \underline{h}_2 \tag{3-12}$$

$$= u(1) + u(2)\, e^{\gamma} \quad . \tag{3-13}$$

Equation (3-13) is a line in the solution space and points on the line correspond to control sequences, u(1) and u(2), that can take $\underline{c}$ into the origin of $\mathcal{C}$ -space. Two such lines are shown in Figure 28, for the initial states c_1 and c_2, $c_2 > c_1 > 0$. The minimum energy points $\underline{u}_1^o$ and $\underline{u}_2^o$, corresponding to c_1 and c_2 respectively, are the points, on the appropriate lines, whose distances from the origin is the least. Since $\underline{u}_1^o$ lies in the square, the amplitude constraints are satisfied; this corresponds to c_1 in M_2. The solution $\underline{u}_2^o$ lies outside the square, with $u^o(2) > 1$. The solution must obviously be moved from $\underline{u}_2^o$ if the saturation constraints are to be satisfied, and must lie on the intersection of the line and the square. The solution with least cost (consider the circles) clearly lies at $\underline{u}^e$ with $u^e(2) = 1$. While this does not prove the postulate for $N > 2$, it does indicate that it is at least reasonable.

An Open Loop Control Procedure

Assuming, therefore, that the postulate is correct, how can it be used to find the optimum amplitude constrained input sequence? The basic philosophy of obtaining the input sequence from Equation (3-10) is best illustrated by means of an example. Consider the plant

$$G_p(s) = \frac{1}{s + \lambda} \quad , \quad \text{with} \quad e^{\lambda T} = e^{\gamma} = 2 \quad , \tag{3-14}$$

and let $\underline{c} = 14$ and $N = 4$. It is desired to find the minimum energy amplitude constrained sequence that takes $\underline{c} = 14$ into the origin in 4 sampling periods.

From Equation (3-1),

$$\underline{h}_1 = 1, \quad \underline{h}_2 = 2, \quad \underline{h}_3 = 4, \quad \underline{h}_4 = 8 \ . \tag{3-15}$$

Equation (3-3) gives

$$u^o(j + 1) = \frac{2^j\ 14}{85} \quad , \tag{3-16}$$

therefore,

$$u^o(4) = 1.32, \quad u^o(3) = 0.66, \quad u^o(2) = 0.33, \quad u^o(1) = 0.16. \tag{3-17}$$

Since $u^o(4) > 1$, the postulate requires $u^e(4) = 1$, which in turn gives

$$\underline{c} = 14 = u(1) + 2u(2) + 4u(3) + 8 \quad . \tag{3-18}$$

The problem now starts again, but with $\underline{c} = 6$ and $N = 3$. Equation (3-3) gives

$$u^o(j + 1) = \frac{2^j\ 6}{21} \quad , \tag{3-19}$$

therefore,

$$u^o(3) = 1.14, \quad u^o(2) = 0.57, \quad u^o(1) = 0.28 \ . \tag{3-20}$$

Setting $u^e(3) = 1$ gives $\underline{c} = 2$ and $N = 2$, giving

$$u^o(2) = 0.8, \quad u^o(1) = 0.4 \ . \tag{3-21}$$

Neither of these exceeds the saturation limit, so the optimum input sequence is,

$$u^e(1) = 0.4, \quad u^e(2) = 0.8, \quad u^e(3) = 1, \quad u^e(4) = 1 \ . \tag{3-22}$$

Figure 29 shows the steps which resulted in Equation (3-22). This simple example has illustrated one general procedure, based on Equation (3-10), that gives the optimum input sequence. This type of procedure generates an open loop control; closed loop control is considered in Chapter V.

III. HIGHER ORDER SYSTEMS

The concepts that have been discussed for first order systems in many cases carry over directly to higher order systems. However, even for second order systems, the concepts are less straightforward, and only in certain cases is it possible to find a reasonably fast method of generating the optimum input sequence. For example, Equation (3-10) does not hold for all second order systems.

The Set M_N

The optimum input sequence, without amplitude constraints, is given by Equations (2-27) and (2-28). These equations are repeated below for the general n-th order system:

$$\underline{a}^o = \begin{bmatrix} u^o(1) \\ \cdot \\ \cdot \\ \cdot \\ u^o(n) \end{bmatrix} = \left[I + HH^t\right]^{-1} \underline{c} \quad , \tag{3-23}$$

$$\underline{b}^o = \begin{bmatrix} u^o(n+1) \\ \cdot \\ \cdot \\ \cdot \\ u^o(N) \end{bmatrix} = H^t \, \underline{a}^o \quad , \tag{3-24}$$

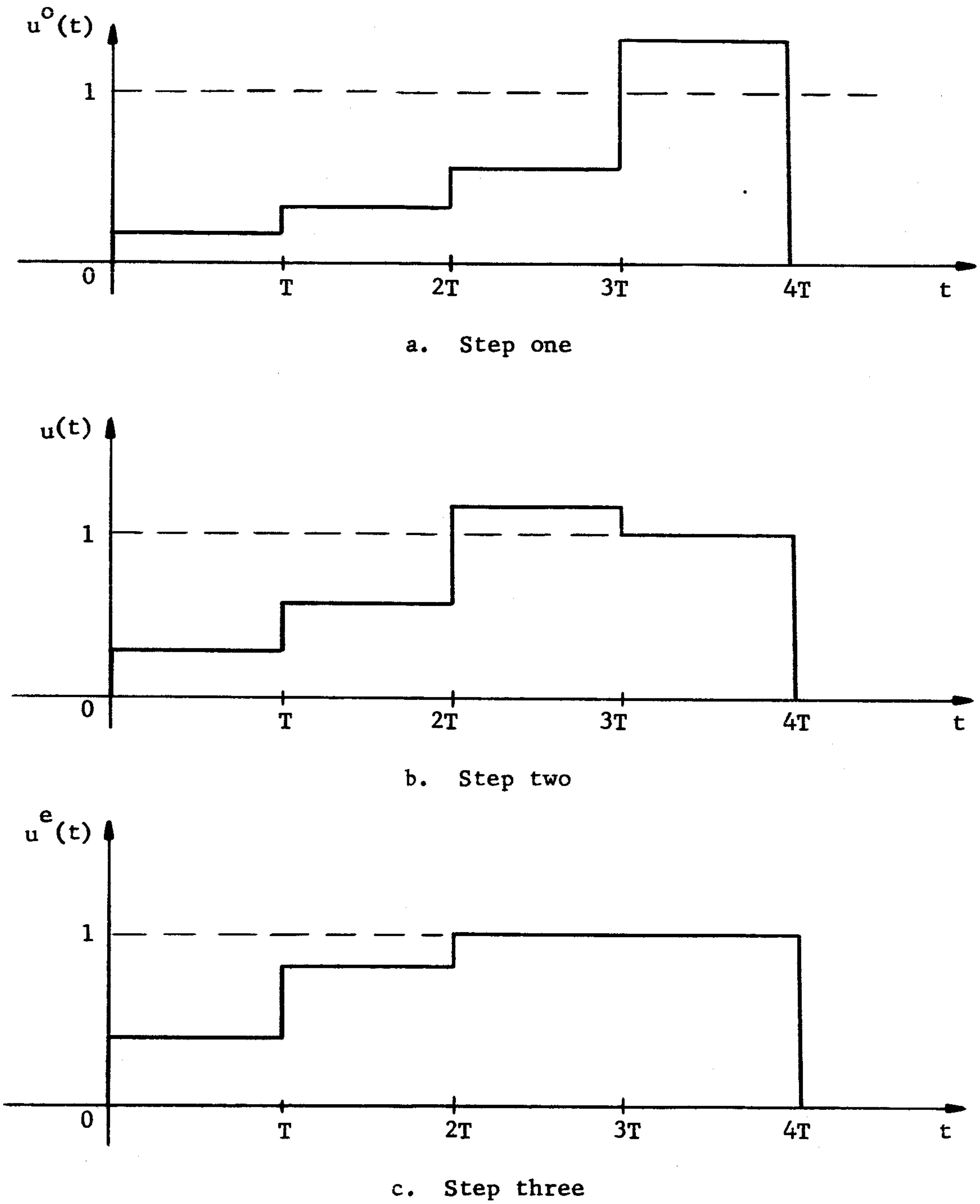


Figure 29. The generation of the constrained minimum energy input sequence for a first order system with N = 4.

where the n x N - n matrix H, given by Equation (2-17) is

$$H = \left[\underline{h}_{n+1}, \underline{h}_{n+2}, \ldots, \underline{h}_N\right] \quad . \tag{3-25}$$

If the initial state $\underline{c}$ is sufficiently close to the origin of $\mathcal{C}$-space (compare Equation (3-6)), none of the input members given by Equations (3-23) and (3-24) will exceed the saturation limit. It is of interest to investigate the set, M_N, of all such initial states, since, if $\underline{c}$ is in M_N, the minimum energy problem is solved without further ado by Equations (3-23) and (3-24). A formal definition of M_N, more general than that given for first order systems, is:

$$M_N = \left(\underline{c} \,\middle|\, \underline{c} = \underline{a} + H\underline{b},\ \underline{b} = H^t\underline{a};\ |u(j)| \leqslant 1,\ j = 1, 2, \ldots, N\right) \quad . \tag{3-26}$$

The set Γ_N, Equation (A-60), can be written as

$$\Gamma_N = \left(\underline{c} \,\middle|\, \underline{c} = \underline{a} + H\underline{b};\ |u(j)| \leqslant 1,\ j = 1, 2, \ldots, N\right) \quad . \tag{3-27}$$

If, for a given settling time of N sampling periods, $\underline{c}$ is Γ_N, solutions to the deadbeat regulator problem exist. The problem of finding which one minimizes the energy subject to the amplitude constraint is solved if $\underline{c}$ is in M_N, since the linear design equations, Equations (3-23) and (3-24), give an input sequence which does not violate the saturation constraint. If $\underline{c}$ is not in M_N, the input sequence, $\underline{u}^o$, will contain at least one member that exceeds the saturation limits. The general properties of M_N are developed next. For the sake of clarity, the discussion is limited to second order systems, but the extension to higher order systems follows without difficulty. The set M_N is most easily imagined as being formed from the lines $u^o(j) = \pm 1$, $j = 1, 2, \ldots,$ N. As an example, Figure 30 shows the sets Γ_3 and M_3 for the plant

$$G_p(s) = \frac{1}{s^2} \quad . \tag{3-28}$$

Some properties of M_N follow directly from Equation (3-26): M_N is convex and symmetric with respect to the origin and is a subset of Γ_N. Figure 31, showing M_N and Γ_N for N = 3 and N = 4 demonstrates a further property of M_N: M_{N+1} does not necessarily include all of the states in M_N. It can, however, be shown for stable and completely controllable plants (compare the similar property of Γ_N in Appendix A, page 243), that as $N \rightarrow \infty$ the set M_N does include all states in $\mathcal{C}$ -space. Since N is a given quantity, however, the saturation problem may not be circumvented by merely increasing N until $\underline{c}$ lies in M_N.

The set M_N may be constructed either by calculating the equations of the lines $u^o(j) = 1$, $j = 1, 2, \ldots, N$, from Equations (3-23) and (3-24) or by the graphical method described in Chapter II. The graphical technique is most conveniently developed in terms of the auxiliary set L_N.

<u>The set L_N.</u> For second order systems, $\mathcal{L}^o$-space, discussed in Chapter II, has coordinates $u^o(1)$ and $u^o(2)$, the members of $\underline{a}^o$. Any point in $\mathcal{L}^o$, corresponding to a particular $\underline{a}^o$, also specifies $\underline{b}^o$ from Equation (3-24). The line $u^o(j) = \pm 1$ is the line

$$\underline{h}_j^t\, \underline{a}^o = \pm 1 \quad . \tag{3-29}$$

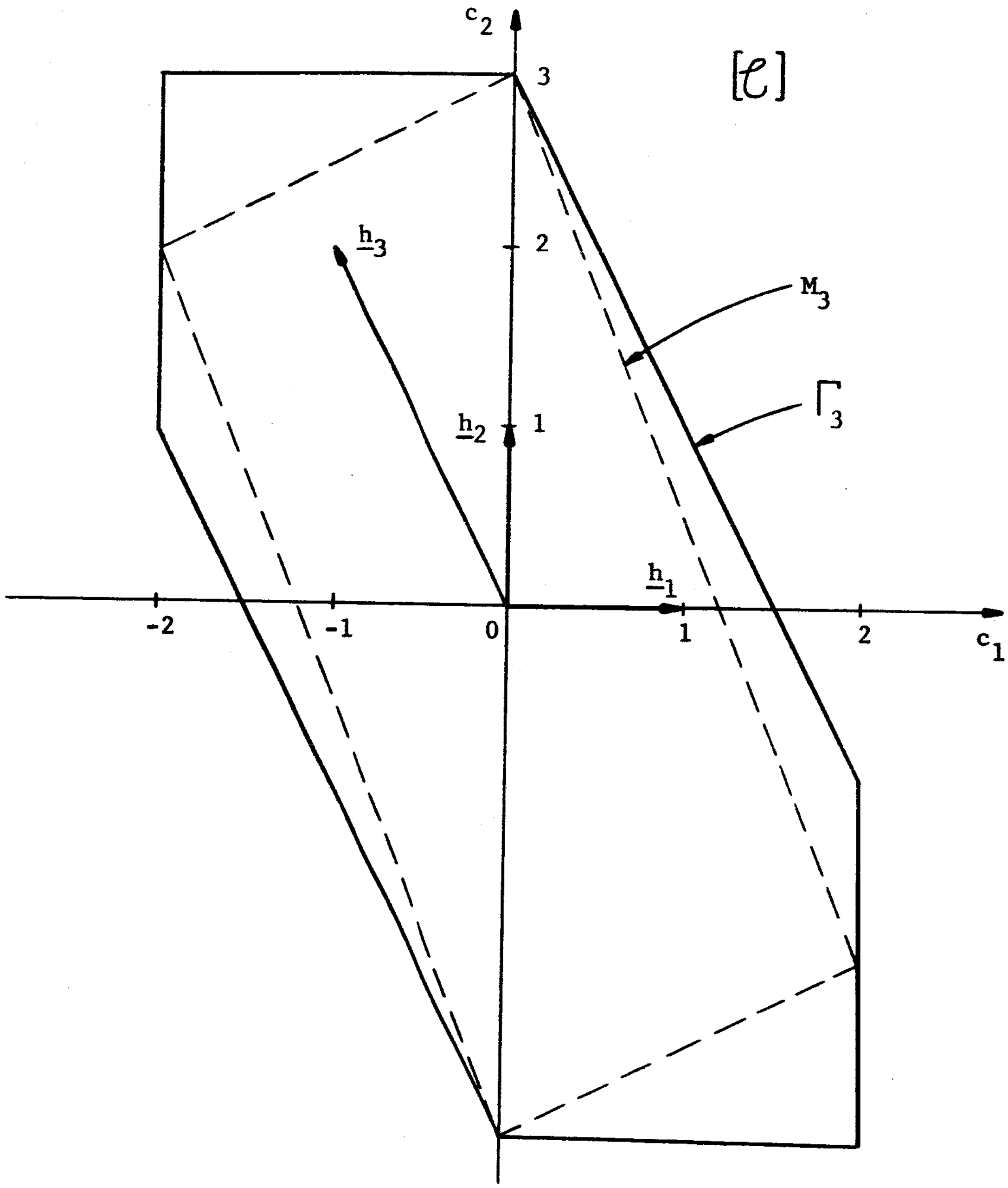


Figure 30. The sets M_3 and Γ_3 for the plant $1/s^2$.

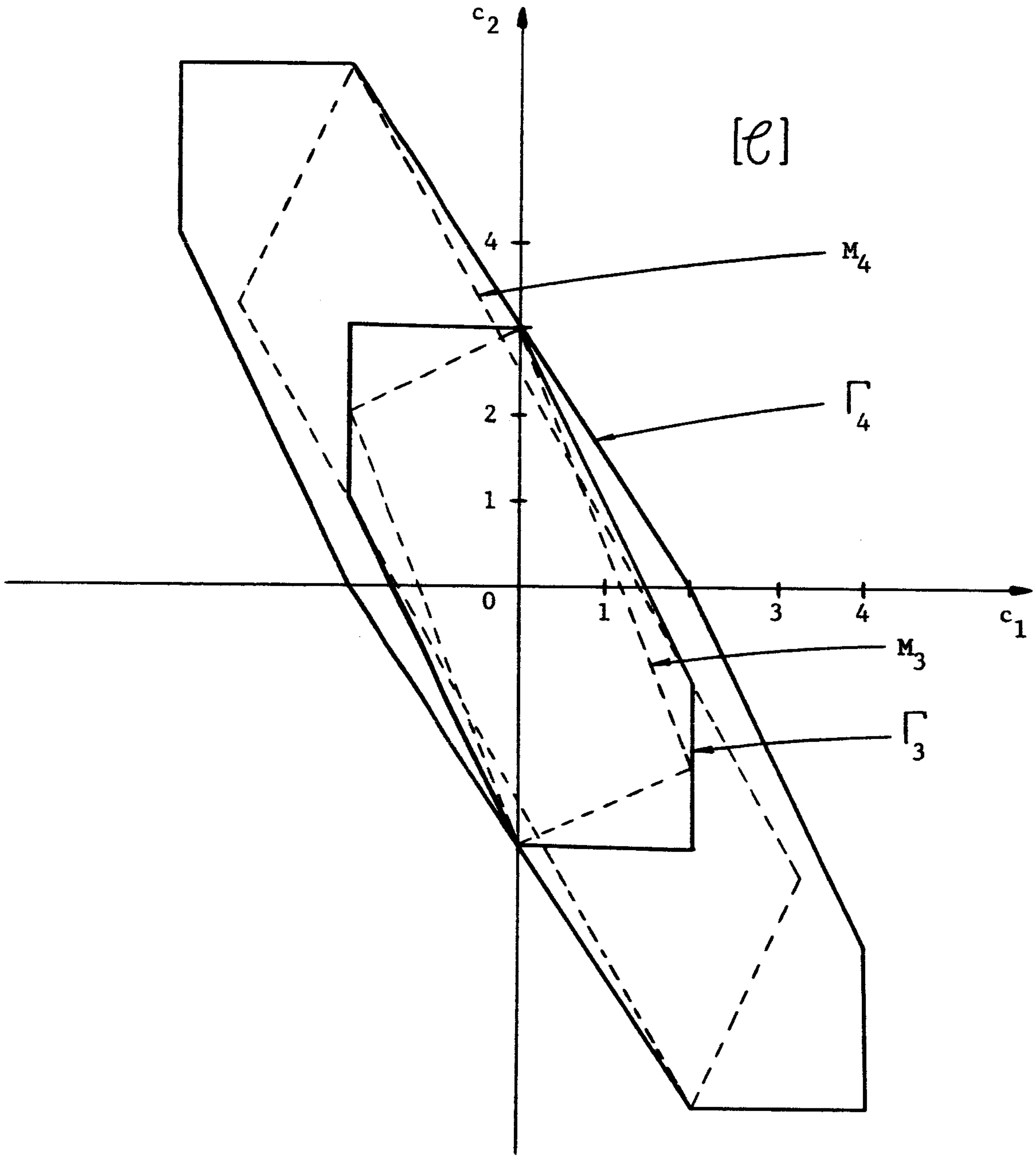


Figure 31. The sets Γ_3, Γ_4, M_3 and M_4 for the plant $1/s^2$.

It will be recalled from Chapter II, that the line $\underline{h}_j^t \, \underline{a}^o = 1$ is normal to the vector $\underline{h}_j$, and intersects $\underline{h}_j$ at a distance $1/\ell_j$ from the origin, where ℓ_j, the length of $\underline{h}_j$, is given by

$$\ell_j = \left[\underline{h}_j^t \, \underline{h}_j\right]^{1/2} \quad . \tag{3-30}$$

Given the invariant vector $\underline{h}_j$, it is, therefore, a straightforward matter to construct the lines of Equation (3-29). Figure 32 shows, for example, the lines $u^o(1) = \pm 1$, $u^o(2) = \pm 1$ and $u^o(3) = \pm 1$ for the plant $1/s^2$. The set L_3 is shown by the cross-hatched area. In general, the set L_N is the set of all $\underline{a}^o$ such that

$$-1 \leqslant \underline{h}_j^t \, \underline{a}^o \leqslant 1, \quad j = 1, 2, \ldots, N \quad . \tag{3-31}$$

Obtaining M_N from L_N. While L_N is defined in $\mathcal{A}^o$-space and M_N is in $\mathcal{C}$-space, they are closely related. Given $\underline{a}^o$ in L_N (and therefore also $\underline{b}^o$) all the members of the input sequence

$$\underline{u}^o = \begin{bmatrix} \underline{a}^o \\ \\ \underline{b}^o \end{bmatrix} \tag{3-32}$$

both satisfy the saturation constraint and minimize the energy cost, and if $\underline{c}$ is in M_N, the input sequence has these same properties. The input sequence and the initial state are related by

$$\underline{c} = \underline{a}^o + \underline{H}\underline{b}^o \quad . \tag{3-33}$$

Therefore, given any point $\underline{a}^o$ in $\mathcal{A}^o$, the input vector $\underline{b}^o$ is fixed, and can be estimated from the lines $\underline{h}_j^t \, \underline{a}^o = 0, \pm 1$; $j = 3, 4, \ldots, N$. The corresponding initial state $\underline{c}$ is given by Equation (3-33), or

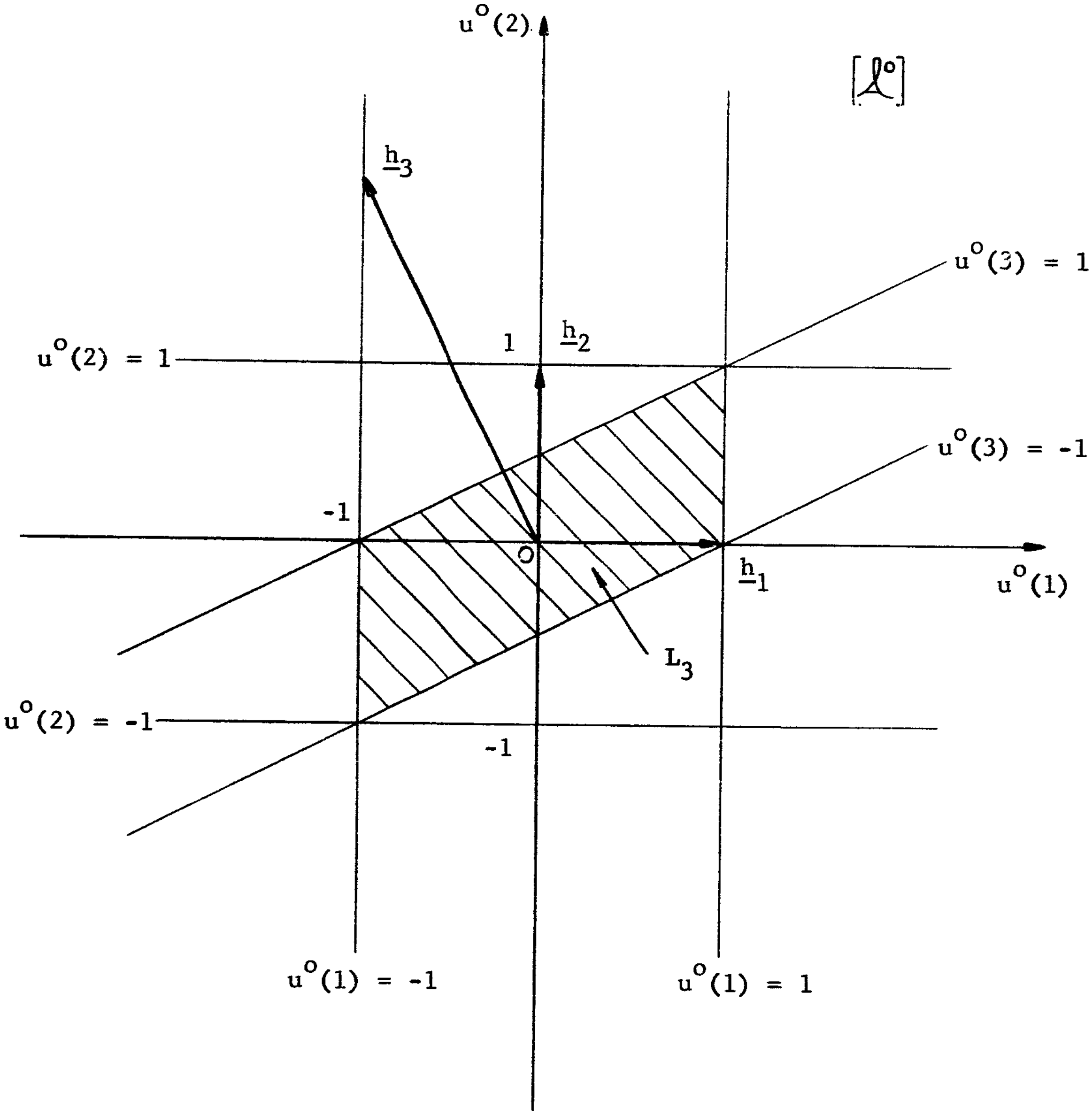


Figure 32. The lines $\underline{h}_j^t\ \underline{a}^o = u^o(j) = \pm 1$, j = 1, 2, 3, for the plant $1/s^2$.

equivalently,

$$\underline{c} = \sum_{j=1}^{N} u^o(j)\ \underline{h}(j) \ . \qquad (3\text{-}34)$$

The set L_N can be drawn directly in $\mathcal{C}$ -space: while generating M_N, the coordinates of $\mathcal{C}$ -space can serve the dual purpose of representing points in both $\mathcal{L}^o$ and $\mathcal{C}$. This device has been discussed already in Chapter II. Figure 33 shows how L_3 is used to generate M_3. The corners of L_N are labelled A, B, C and D; corresponding points on M_N are labelled A', B', C' and D'.

The Problem of Saturation

Having found the set M_N and its size relative to Γ_N, the likelihood that $\underline{c}$ lies in M_N becomes evident, see for example Figures 30, page 82, and 31, page 83. However, just as for first order systems, the question arises: for $\underline{c}$ in Γ_N, what is the optimum input sequence if $\underline{c}$ does not lie in M_N? Unfortunately, there is no simple general postulate, such as Equation (3-10), which can be used to obtain the optimum constrained input sequence. Stubberud and Swiger (36) gave a method which purported to indicate which members of the input sequence satisfied

$$u^e(j) = \text{sgn. } u^o(j) \ . \qquad (3\text{-}35)$$

However, the method is based upon a theorem (36, page 405) which, as will be shown later in this chapter, breaks down in certain cases. The analogy, which proved useful for first order systems, and which might

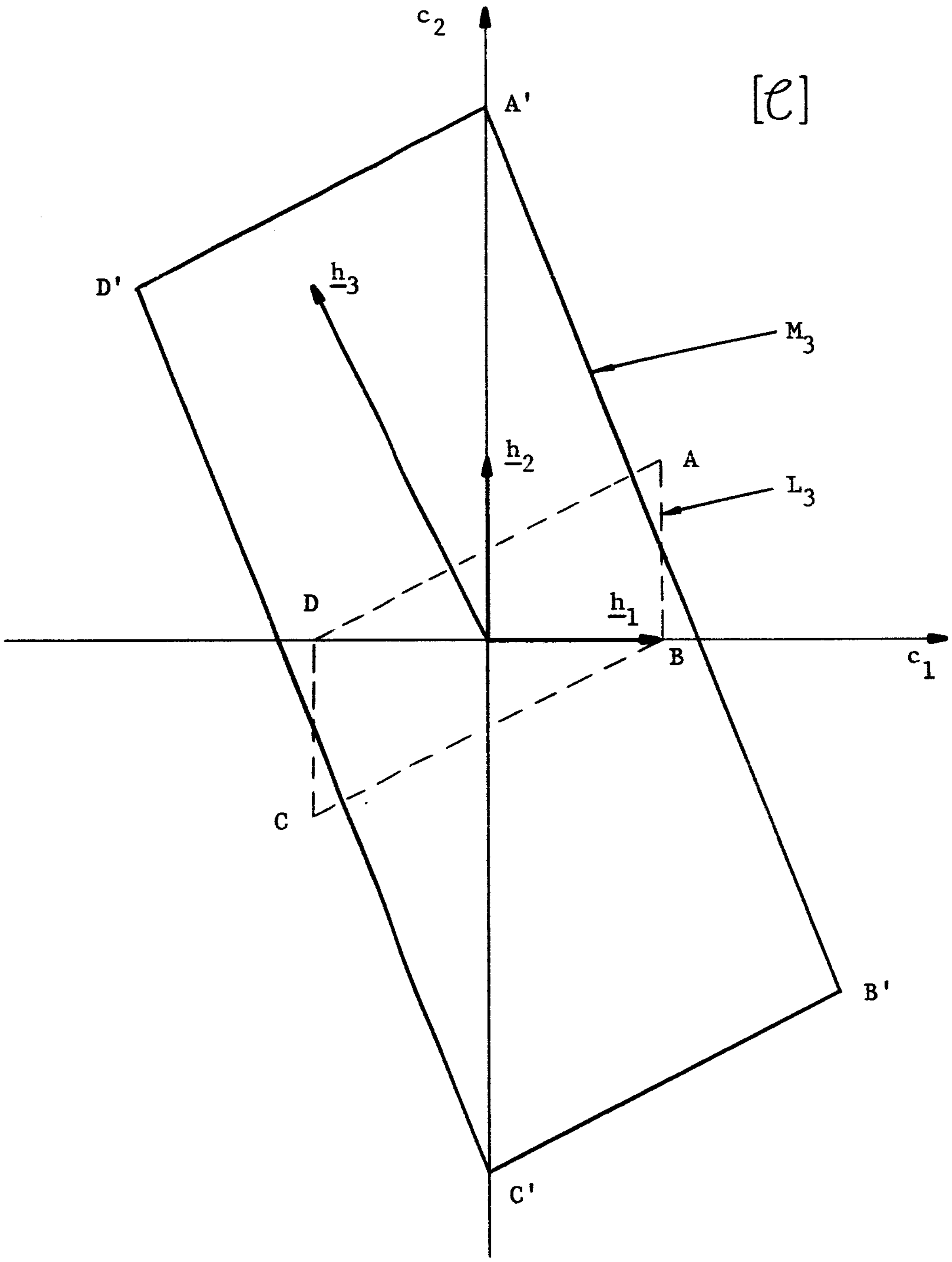


Figure 33. The generation of the set M_N from the set L_N for N = 3.

lead one to expect that Equation (3-35) is true for every j for which $|u^o(j)| > 1$, $j = 1, 2, \ldots, N$, does not reveal the subtleties of the saturation problem, even for second order systems. The results obtained by Stubberud and Swiger were developed in the solution space. The drawback is that if $N > 3$, visualization of the problem becomes impossible; even with $N = 3$ it is not easy. By equivalently partitioning the solution space into two spaces, one n-dimensional, the other (N-n)-dimensional, the saturation problem can be examined more readily.

The Correction Space

Suppose, having calculated the unconstrained optimum input sequence, $\underline{u}^o$, from Equations (3-23) and (3-24), it is found that one or more of the members of $\underline{u}^o$ violate the saturation constraint. Consider the effect of adding to $\underline{u}^o$ an N x 1 correction vector $\underline{\delta}$. There results a new input sequence, $\underline{u}$, given by

$$\underline{u} = \underline{u}^o + \underline{\delta} \quad . \tag{3-36}$$

For $\underline{u}$ to be considered as a candidate for the optimum constrained input sequence, $\underline{u}^e$, $\underline{u}$ must satisfy both Equation (1-10),

$$C\,\underline{u} = \underline{x}(0) \quad , \tag{3-37}$$

and

$$|u(j)| \leqslant 1, \quad j = 1, 2, \ldots, N \quad . \tag{3-38}$$

Substituting Equation (3-38) into (3-37) gives

$$C\,\underline{u}^o + C\,\underline{\delta} = \underline{x}(0) \quad , \tag{3-39}$$

and since $\underline{u} = \underline{u}^o$ already satisfies Equation (3-37), Equation (3-39) becomes

$$C\underline{\delta} = 0 \quad . \tag{3-40}$$

Equation (3-36) gives

$$u(j) = u^{o}(j) + \delta(j) \quad , \tag{3-41}$$

where $\delta(j)$ is the j-th member of $\underline{\delta}$. In order that u(j) satisfy Equation (3-38), $\delta(j)$ must satisfy,

$$-1 - u^{o}(j) \leqslant \delta(j) \leqslant 1 - u^{o}(j) \ , \quad j = 1, 2, \ldots, N \ . \tag{3-42}$$

Equations (3-40) and (3-42) are constraints that must be satisfied by $\underline{\delta}$ so that $\underline{u} = \underline{u}^{o} + \underline{\delta}$ takes $\underline{x}(0)$ to the origin and satisfies the saturation constraint. From these allowable corrections, $\underline{\delta}$, will be selected the one that gives $E = \underline{u}^{t}\ \underline{u}$ a minimum value.

The energy taken by the control sequence $\underline{u}$ is

$$E = \underline{u}^{t}\ \underline{u} = \underline{u}^{ot}\ \underline{u}^{o} + 2\underline{u}^{ot}\underline{\delta} + \underline{\delta}^{t}\underline{\delta} \quad . \tag{3-43}$$

With the use of Equation (2-37),

$$E = E^{o} + 2\left[C^{t}(CC^{t})^{-1}\ \underline{x}(0)\right]^{t}\underline{\delta} + \underline{\delta}^{t}\underline{\delta} \quad . \tag{3-44}$$

Therefore,

$$E - E^{o} = 2\underline{x}(0)^{t}(CC^{t})^{-1}\ C\underline{\delta} + \underline{\delta}^{t}\underline{\delta} \quad , \tag{3-45}$$

and since $C\underline{\delta} = 0$,

$$E - E^{o} = \Delta E = \underline{\delta}^{t}\underline{\delta} \quad . \tag{3-46}$$

Equation (3-46) says that if a correction $\underline{\delta}$ is added to the unconstrained minimum energy sequence, $\underline{u}^{o}$, the resulting input sequence, $\underline{u}$, requires an extra amount of energy, $\Delta E = \underline{\delta}^{t}\underline{\delta}$. If the members,

$\underline{\delta}(j)$, j = 1, 2, ..., N, of the correction vector are made the coordinates of an N-dimensional correction space, the $\underline{\delta}$ that satisfy Equation (3-41) lie in an N-dimensional hypercube centered on $\underline{\delta} = -\underline{u}^o$. The constraint of Equation (3-40) is an (N-n)-dimensional hyperplane through the origin of the correction space. Figure 34 shows the correction space for a first order system with N = 2; compare this with the solution space in Figure 28, page 75. Note that Figure 34 shows no intersection between the square and the line, and there is, therefore, no solution to this amplitude constrained regulator problem; $\underline{c}$ does not lie in Γ_2. At this point there seems little advantage in the correction space; visualization of the problem for $N > 2$ is again difficult or impossible. However, by partitioning the correction vector $\underline{\delta}$ the problem can be visualized for N = 4 or even 5.

The Partitioned Solution Space

Let the correction vector $\underline{\delta}$ be partitioned so that

$$\underline{\delta} = \begin{bmatrix} \underline{\alpha} \\ \underline{\beta} \end{bmatrix} \qquad (3\text{-}47)$$

where $\underline{\alpha}$ is an n-vector, corresponding to a correction to $\underline{a}^o$, and $\underline{\beta}$ is an (N-n)-vector, corresponding to a correction to $\underline{b}^o$: $\alpha(j) = \delta(j)$, j = 1, 2, ..., n, and $\beta(j-n) = \delta(j)$, j = n+1, ..., N are the components of $\underline{\alpha}$ and $\underline{\beta}$. Equation (3-40) becomes

$$R\,\underline{\alpha} + Q\,\underline{\beta} = 0 \quad , \qquad (3\text{-}48)$$

which, on multiplying by R^{-1}, gives

$$\underline{\alpha} + H\,\underline{\beta} = 0 \quad , \qquad (3\text{-}49)$$

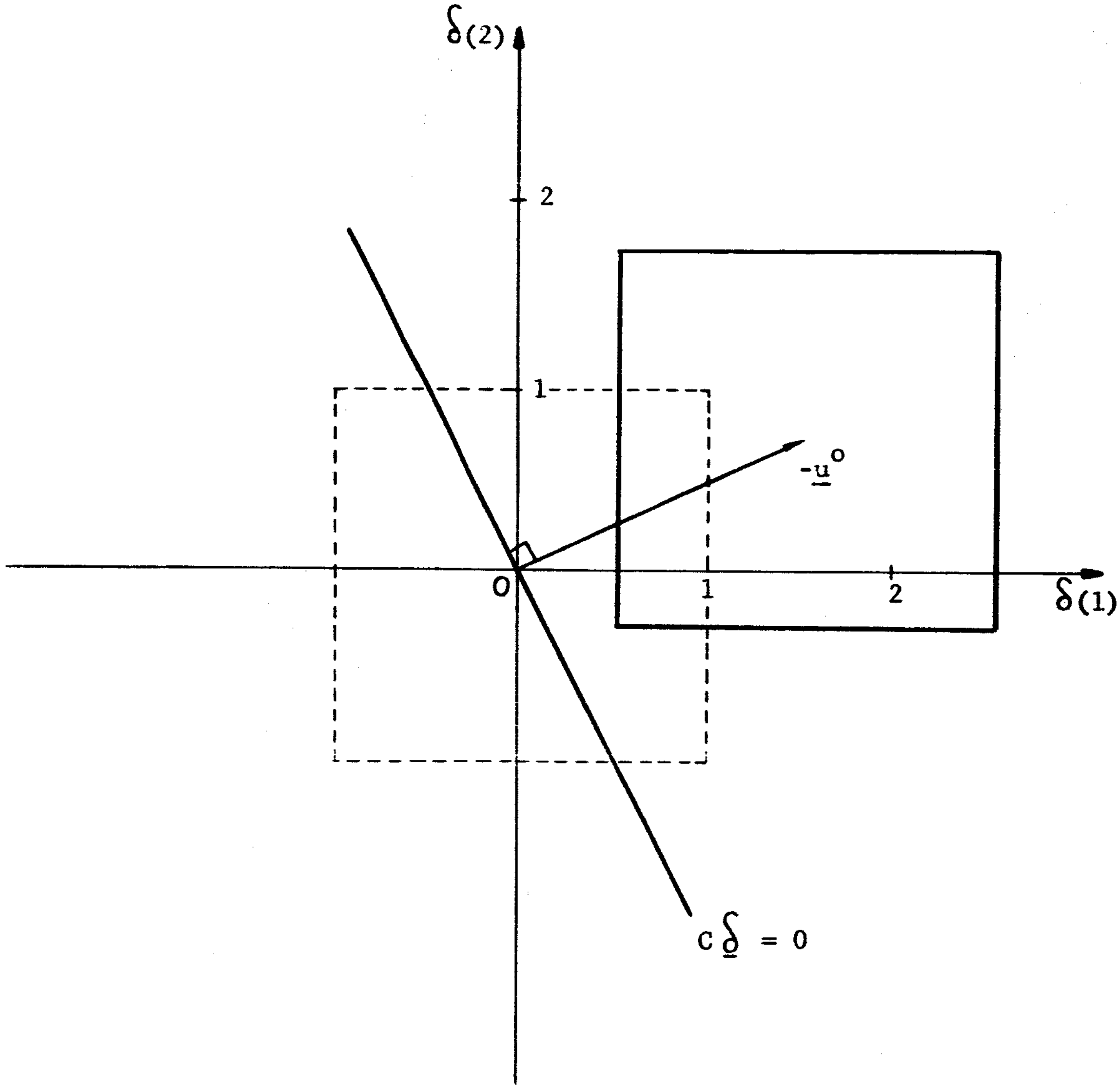


Figure 34. The correction space for N = 2.

and the energy correction, ΔE, is therefore

$$\Delta E = \underline{\alpha}^t\,\underline{\alpha} + \underline{\beta}^t\,\underline{\beta} = \underline{\beta}^t\left[I + H^t\,H\right]\underline{\beta} \quad . \tag{3-50}$$

The minimum energy regulator with amplitude constrained input sequence has been transformed to the problem: minimize

$$\Delta E = \underline{\beta}^t\left[I + H^t\,H\right]\underline{\beta} \tag{3-51}$$

subject to

$$-1 - u^o(j) \leqslant \alpha(j) \leqslant 1 - u^o(j)\ ,\ j = 1, 2, \ldots, n, \tag{3-52}$$

$$-1 - u^o(n+j) \leqslant \beta(j) \leqslant 1 - u^o(n+j),\ j = 1, 2, \ldots, N-n, \tag{3-53}$$

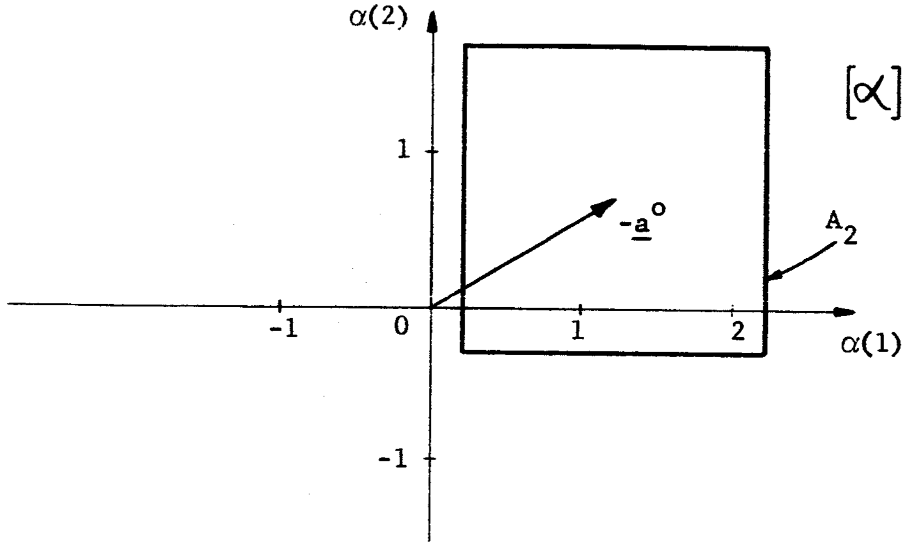
and

$$H\,\underline{\beta} = -\,\underline{\alpha} \quad . \tag{3-54}$$

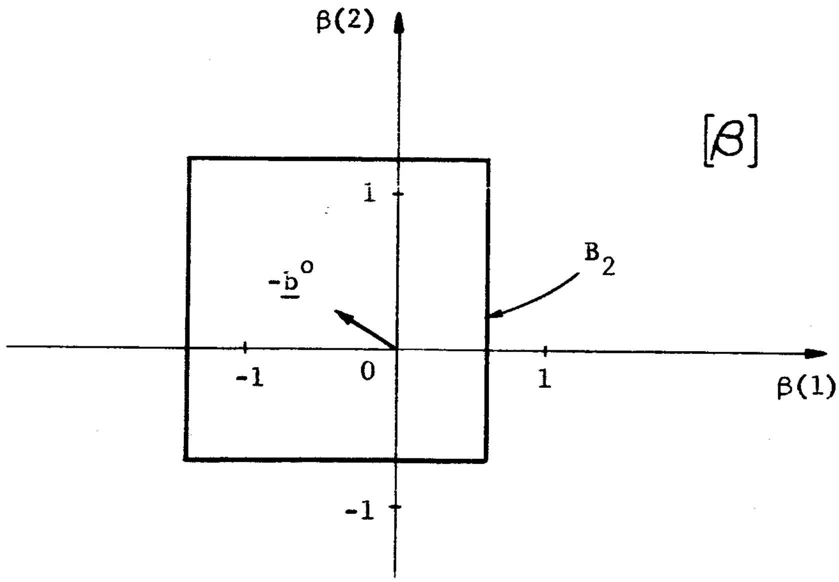
Let α-space be the n-dimensional space with coordinates $\alpha(1), \ldots, \alpha(n)$, and let β-space have coordinates $\beta(1), \ldots, \beta(N-n)$. The direct sum of α-space and β-space is, of course, the correction space. Denote by A_n the set of $\underline{\alpha}$ that satisfy Equation (3-52), and by B_{N-n} the set of $\underline{\beta}$ that satisfy Equation (3-53). These sets are shown in Figure 35 for a second order system (n = 2) with N = 4. The direct sum of A_n and B_{N-n} is just the N-dimensional hypercube of $\underline{\delta}$ that satisfy Equation (3-42).

Now, assuming $N \geqslant 4$, by means of Equation (3-54), the set A_n can be transformed into β-space. Let this map of A_n be called A_n'. For second order systems, Equation (3-54) gives

$$\begin{bmatrix} h_{31} & h_{41} & \cdots & h_{N1} \\ h_{32} & h_{42} & \cdots & h_{N2} \end{bmatrix} \begin{bmatrix} \beta(1) \\ \cdot \\ \cdot \\ \cdot \\ \beta(N-2) \end{bmatrix} = \begin{bmatrix} -\alpha(1) \\ -\alpha(2) \end{bmatrix} \quad . \tag{3-55}$$



a. The set A_2



b. The set B_2

Figure 35. The sets A_n and B_{N-2} for n = 2 and N = 4.

Equation (3-55) gives two hyperplanes in β-space. If in general the $1 \times N$ matrix $\underline{p}_j$, $j = 1, 2, \ldots, n$, is defined as

$$\underline{p}_j = \left[h_{n+1\ j}, h_{n+2\ j}, \ldots, h_{N\ j}\right] \quad , \tag{3-56}$$

the equations of the two hyperplanes, Equation (3-55), may be written as

$$\underline{p}_1\ \underline{\beta} + \alpha(1) = 0 \quad , \tag{3-57}$$

$$\underline{p}_2\ \underline{\beta} + \alpha(2) = 0 \quad . \tag{3-58}$$

The row vectors $\underline{p}_j$, $j = 1, 2, \ldots, n$, are simply the n rows of the matrix H. The set A_2' is, therefore, the set of points, $\underline{\beta}$, that are generated by Equations (3-58) and (3-59) for all $\underline{\alpha}$ in A_2.

The intersection of A_n' and B_{N-n} defines a set U_{N-n}. Corrections, $\underline{\beta}$, lying in U_{N-n} satisfy Equation (3-53); the corresponding $\underline{\alpha}$, given by Equation (3-54), satisfy Equation (3-52). Therefore, any $\underline{\beta}$ lying in U_{N-n} is a possible choice to minimize the correction energy given by Equation (3-51). Equation (3-51) describes ΔE as a positive definite quadratic form. In a two dimensional β-space, for a constant ΔE, Equation (3-51) in gneral describes an ellipse; in three dimensions an ellipsoid, and so on.

<u>Example</u>. To illustrate how A_n' and the ellipsoid depend on the matrix H, consider the second order plant,

$$G_p(s) = \frac{1}{(s + a + jb)(s + a - jb)} \quad , \tag{3-59}$$

and allow a settling time of four sampling periods ($n = 2$, $N = 4$). The invariant vectors, $\underline{h}_j$, $j = 3, 4$, for this plant are obtained from

Table I, Appendix A, page 252:

$$\underline{h}_3 = \begin{bmatrix} e^{2aT} \\ \\ 2e^{aT}\cos bT \end{bmatrix} , \quad \underline{h}_4 = \begin{bmatrix} -2e^{3aT}\cos bT \\ \\ e^{2aT}(4\cos^2 bT - 1) \end{bmatrix} . \tag{3-60}$$

If a = 3.465, b = 14.83 and the sampling period is chosen as T = 0.1, these invariant vectors become,

$$\underline{h}_3 = \begin{bmatrix} -2 \\ \\ 0.25 \end{bmatrix} , \quad \underline{h}_4 = \begin{bmatrix} -0.5 \\ \\ -1.9375 \end{bmatrix} . \tag{3-61}$$

From Equation (3-25),

$$H = \begin{bmatrix} -2 & -0.5 \\ 0.25 & -1.9375 \end{bmatrix} . \tag{3-62}$$

Consider the generation of the set A_2'. From Equation (3-56),

$$\underline{p}_1 = \begin{bmatrix} -2, & -0.5 \end{bmatrix} , \quad \underline{p}_2 = \begin{bmatrix} 0.25, & -1.9375 \end{bmatrix} . \tag{3-63}$$

For the moment assume $\underline{a}^o = 0$. Equations (3-57) and (3-58) are used to construct the set A_2', shown in Figure 36 as the dashed parallelogram. Note that if $\underline{a}^o \neq 0$, the set A_2' would not be centered on the origin; the case $\underline{a}^o = 0$ corresponds to the unrealistic case of $\underline{c} = 0$. Furthermore, as long as both $|u^o(1)| \leqslant 1$ and $|u^o(2)| \leqslant 1$, the set A_2' will contain the origin of β-space. Now consider the shape of the ellipse. The positive definite matrix in Equation (3-51) is

$$\begin{bmatrix} I + H^t H \end{bmatrix} = \begin{bmatrix} 5.0625 & 0.515625 \\ \\ 0.515625 & 5.00390625 \end{bmatrix} . \tag{3-64}$$

The eigenvalues, θ_1 and θ_2, of this matrix are

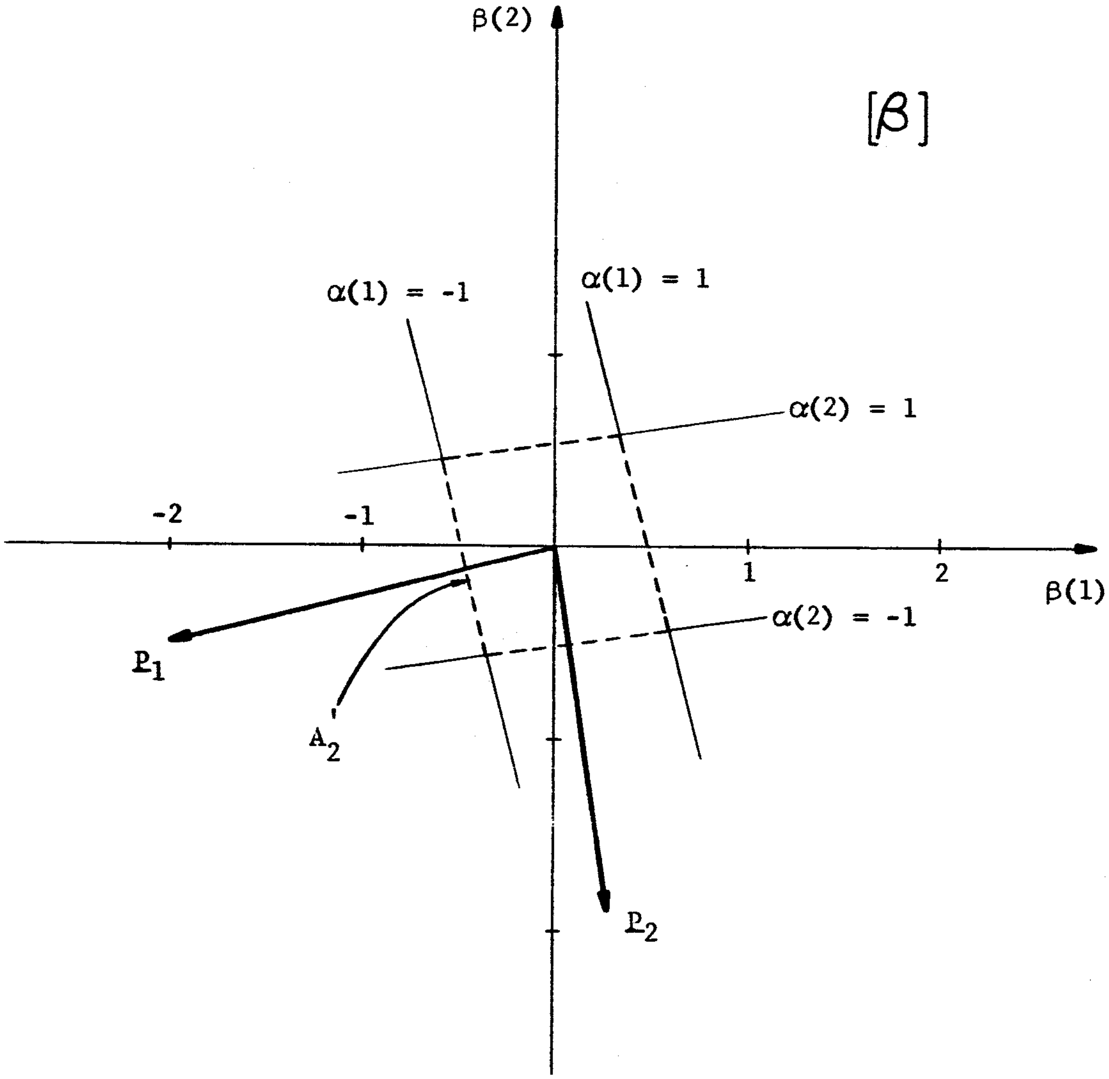


Figure 36. The set A_2' in a two-dimensional β -space.

$$\theta_1 = 4.5167, \qquad \theta_2 = 5.5497 \tag{3-65}$$

and the corresponding eigenvectors, $\underline{\omega}_1$ and $\underline{\omega}_2$ respectively are

$$\underline{\omega}_1 = \begin{bmatrix} 1 \\ -1.058 \end{bmatrix}, \qquad \underline{\omega}_2 = \begin{bmatrix} 1 \\ 0.945 \end{bmatrix}. \tag{3-66}$$

The eigenvectors correspond to the major and minor axes of the ellipse; the <u>major</u> axis lies along the eigenvector ($\underline{\omega}_1$) formed from the <u>smaller</u> eigenvalue (θ_1), and the ratio of the length of the major axis to the length of the minor axis is

$$\left[\frac{\theta_2}{\theta_1}\right]^{1/2} . \tag{3-67}$$

As ΔE increases, the ellipse becomes larger. Figure 37 shows ellipses corresponding to ΔE_1 and ΔE_2, $\Delta E_2 > \Delta E_1$.

<u>The set U_{N-n}.</u> Having considered the special case of a second order system in a two dimensional β-space to illustrate the generation of A_2' and the ellipse of Equation (3-51), it remains to consider the set U_{N-n}. Since the set B_{N-n} is a convex set, and A_n' is a convex set, their intersection, the set U_{N-n}, is a convex set. The faces of the set U_{N-n} in general consist of the faces of the hypercube B_{N-n} and/or the 2n hyperplanes

$$\underline{p}_j\, \underline{\beta} + \alpha(j) = 0,\ \alpha(j) = 1 - u^o(j),\ \alpha(j) = -1 - u^o(j),\ j = 1, \ldots, n. \tag{3-67}$$

If the initial state $\underline{c}$ is not in Γ_N, the sets A_n' and B_{N-n} are disjoint;

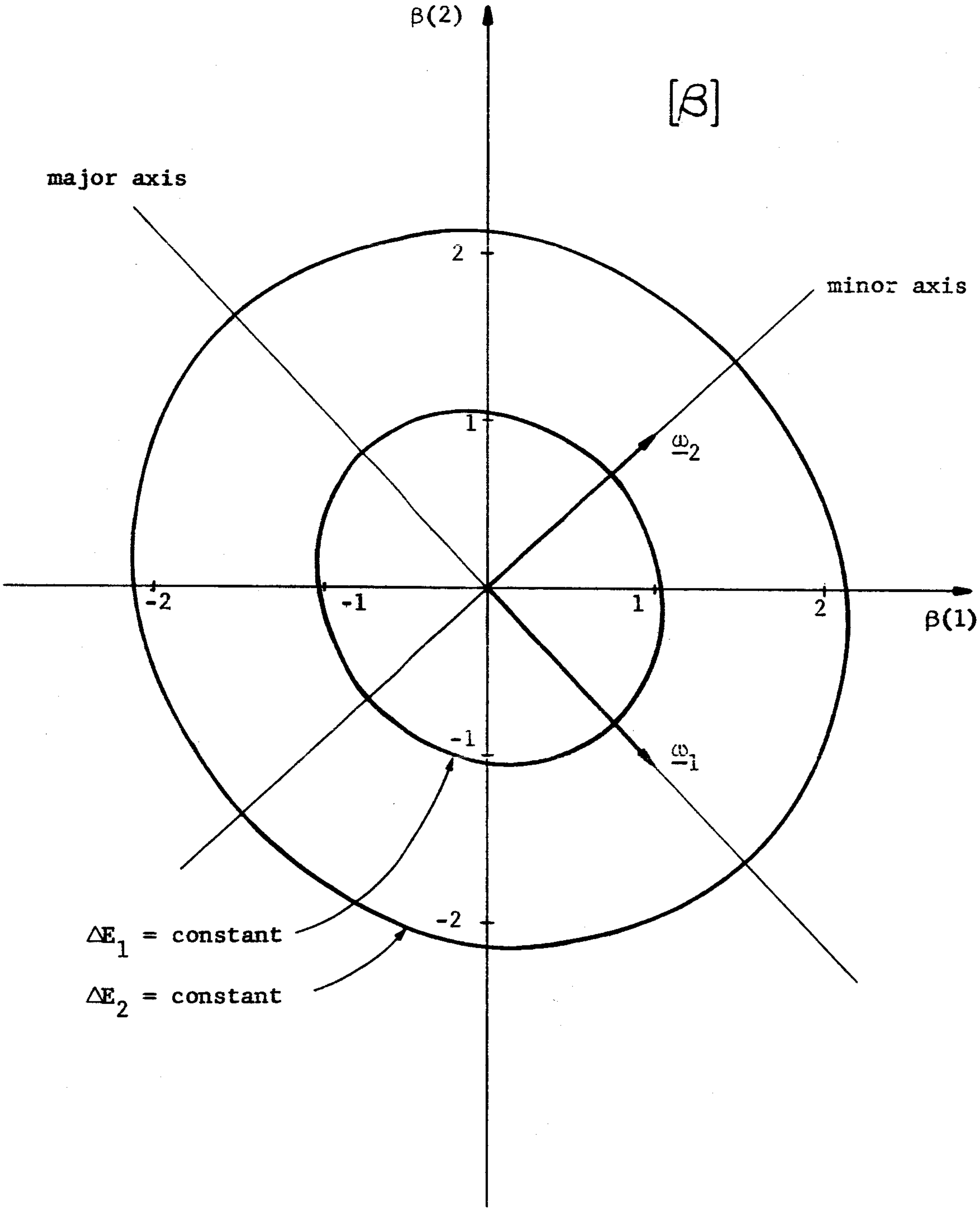


Figure 37. Ellipses of constant correction energy in β-space.

i.e., they contain no points in common, and U_{N-n} is the null set. If $\underline{c}$ is in Γ_N, U_{N-n} does contain at least one point. Any $\underline{\beta}$ in U_{N-n} gives, with Equations (3-47) and (3-54), a correction $\underline{\delta}$, which gives, from Equation (3-36), in turn a control sequence $\underline{u}$, satisfying the saturation constraint and taking $\underline{c}$ to the origin of $\mathcal{C}$-space. If the initial state $\underline{c}$ is in M_N, clearly no correction is necessary; in this case, U_{N-n} contains the origin of β-space. If the initial state is not in M_N, not all of the members of $\underline{u}^o$ lie within the saturation limits, and it becomes necessary to add a correction $\underline{\delta}$ to $\underline{u}^o$. In this case, U_{N-n} does not contain the origin of β-space.

<u>The minimum energy problem reformulated in β-space</u>. The minimum energy problem with input saturation can now be considered in the following manner. For a given settling time, N sampling periods, and a given initial state $\underline{c}$, $\underline{c}$ in Γ_N, the solution to the minimum energy problem <u>without</u> input saturation is $\underline{u}^o$, and is obtained from Equations (3-23) and (3-24). If $|u^o(j)| \leqslant 1$, $j = 1, 2, \ldots, N$, the problem is solved. In such a case, U_{N-n} containing the origin of β-space, no correction is necessary. If however, one or more of the input members $u^o(j)$ exceeds the saturation limits, $|u^o(j)| > 1$, the set U_{N-n} does not contain the origin, and a correction $\underline{\beta}$ is, therefore, required. The energy required by the corrected sequence, $\underline{u} = \underline{u}^o + \underline{\delta}$, is, from Equation (3-46), given by $E = E^o + \Delta E$. This energy is minimized when ΔE is minimized. As ΔE increases, the (N-n)-dimensional ellipsoidal surface of Equation (3-51) moves outward from the origin

(compare Figure 37). The permissible correction, $\underline{\beta}$, must lie in the set U_{N-n}. Therefore, if ΔE is smoothly increased from $\Delta E = 0$ until the surface first touches the set U_{N-n}, this first point of contact is the minimum energy correction. Let this point be $\underline{\beta}^e$. Then the solution to the minimum energy problem with input saturation is

$$\underline{u}^e = \underline{u}^o + \underline{\delta}^e \tag{3-68}$$

where

$$\underline{\delta}^e = \begin{bmatrix} \underline{\alpha}^e \\ \\ \underline{\beta}^e \end{bmatrix} \quad ; \quad \underline{\alpha}^e = - H \underline{\beta}^e \quad ; \tag{3-69}$$

and the minimum energy is

$$E^e = E^o + \underline{\delta}^{e^t} \underline{\delta}^e \quad . \tag{3-70}$$

Of course, the problem still requires a numerical solution. The problem of finding the point $\underline{\beta}^e$ is, in general, not a trivial matter. Before pursuing this problem any further, it is necessary to identify the various faces of U_{N-n}.

In β-space let the (N-n)-dimensional hyperplane

$$\beta(j) = - u^o(n + j) + 1 \ , \ j = 1, 2, \ldots, N-n \tag{3-71}$$

be denoted W_{n+j}, and let the (N-n)-dimensional hyperplane

$$\beta(j) = - u^o(n + j) - 1 \ , \quad j = 1, 2, \ldots, N-n, \tag{3-72}$$

be denoted $W_{-(n+j)}$. These hyperplanes form the hypercube B_{N-n}. The boundaries of the set A_n' are the 2n hyperplanes given by Equation (3-67). Let the (N-n)-dimensional hyperplane,

$$\underline{p}_j\,\underline{\beta} - u^o(j) + 1 = 0\ ,\quad j = 1, 2, \ldots, n\ , \tag{3-73}$$

be denoted W_j, and let W_{-j} denote the hyperplane

$$\underline{p}_j\,\underline{\beta} - u^o(j) - 1 = 0\ ,\quad j = 1, 2, \ldots, n\ . \tag{3-74}$$

These 2N hyperplanes, W_j, W_{-j}, $j = 1, 2, \ldots, N$, define the set U_{N-n}. Depending on the initial state $\underline{c}$, not all the hyperplanes are necessarily faces of U_{N-n}.

General approaches to the problem of finding $\underline{\beta}^e$. From Equations (3-41) and (3-47), if $\underline{\beta}$ lies in W_j, $u(j) = 1$. Similarly, if $\underline{\beta}$ lies in W_{-j}, $u(j) = -1$. The optimum correction, $\underline{\beta}^e$, lying on the boundary of U_{N-n}, must lie in one or more of the hyperplanes W_j, W_{-j}, $j = 1, 2, \ldots, N$. Therefore, the minimum energy problem amounts to finding which members of the optimum input sequence, $\underline{u}^e$, equal the saturation limit. As an example, Figure 38 shows the sets B_2, A_2' and U_2 for a typical second order system. The set A_2' is shown by the dashed parallelogram, and U_2 is the cross-hatched area. The boundaries of A_2' and B_2 are labelled by their corresponding lines W_j, W_{-j}, $j = 1, 2, 3, 4$, and the optimum correction, $\underline{\beta}^e$, is shown at the intersection of W_4 and W_2.

The general problem of finding $\underline{\beta}^e$ obviously depends on the shape and position of the set U_{N-n} relative to the surfaces of Equation (3-51), and some general observations can be made. If none of the members of $\underline{a}^o$, $u^o(1)$, $u^o(2)$, $\ldots$, $u^o(n)$, exceed the saturation limit, A_n' contains the origin of β-space, and $\underline{\beta}^e$ must lie on the boundary of B_{N-n} and not in its interior. Therefore, at least one of the members of $\underline{b}^e$, $u^e(n+j)$,

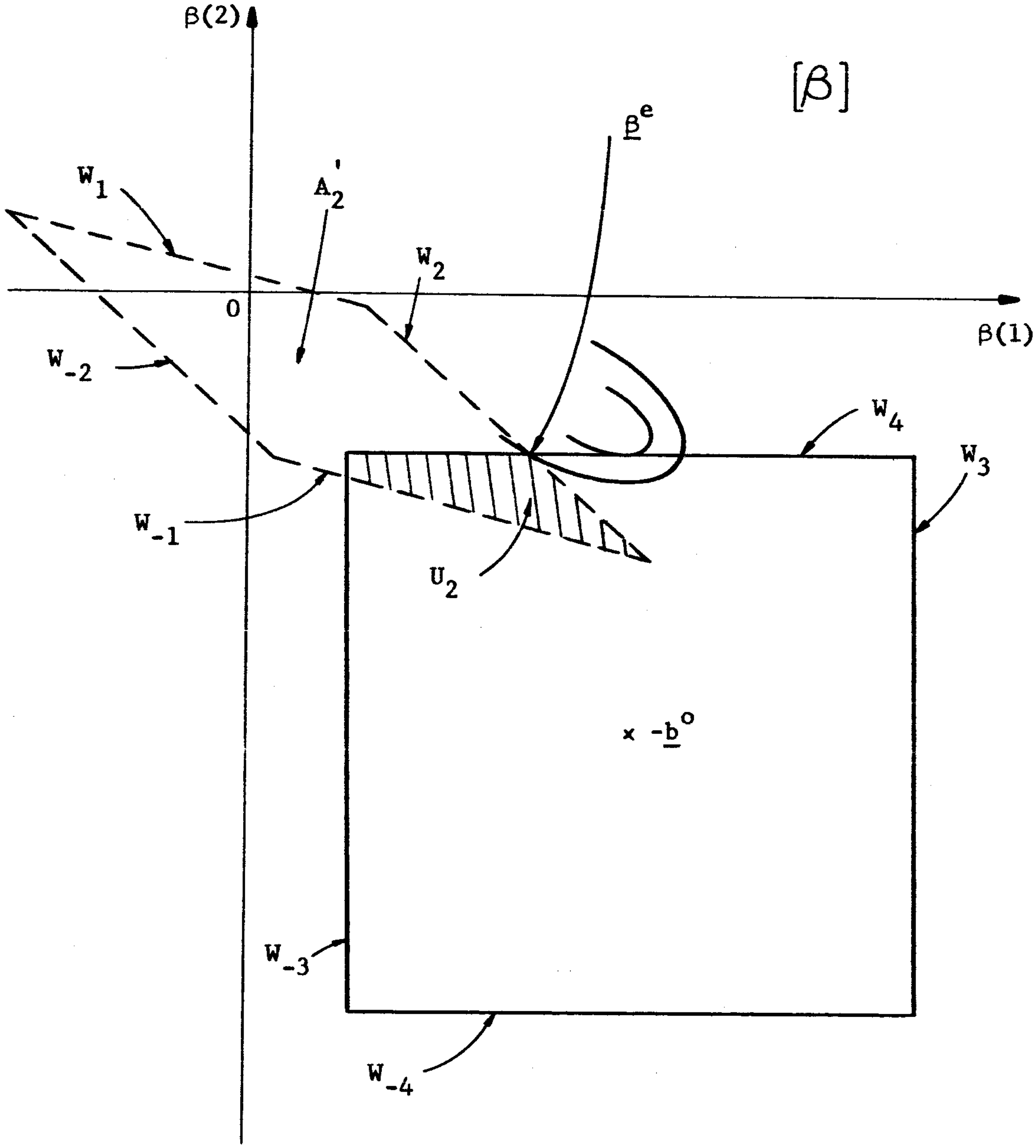


Figure 38. The sets B_2 and A_2' forming the set U_2, with the optimum correction $\underline{\beta}^e$.

$j = 1, 2, \ldots, N-n$, must be equal to the saturation limit. Similarly, if none of the members, $u^o(n+1), \ldots, u^o(N)$, of $\underline{b}^o$ exceeds the saturation limit, B_{N-n} contains the origin and $\underline{\beta}^e$ lies on at least one of the W_j, W_{-j}, $j = 1, 2, \ldots, n$. Therefore, at least one $u^e(j)$, $j = 1, 2, \ldots, n$, equals the saturation limit. If only one of the members of $\underline{u}^o$ exceeds the saturation limit, the following theorem is applicable.

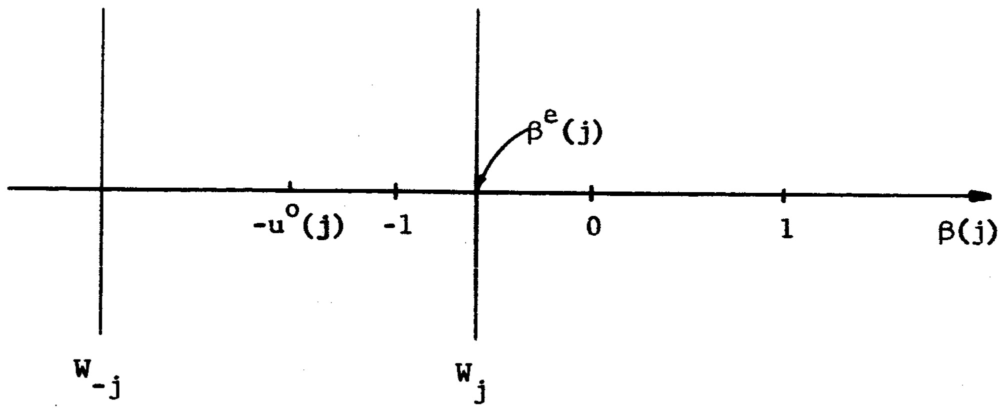
Theorem 2. If $\underline{c}$ is in Γ_N but not in M_N, and if only one of the members, let it be the j-th member, of $\underline{u}^o$ exceeds the saturation limits, then

$$u^e(j) = \text{sgn. } u^o(j) \quad . \tag{3-75}$$

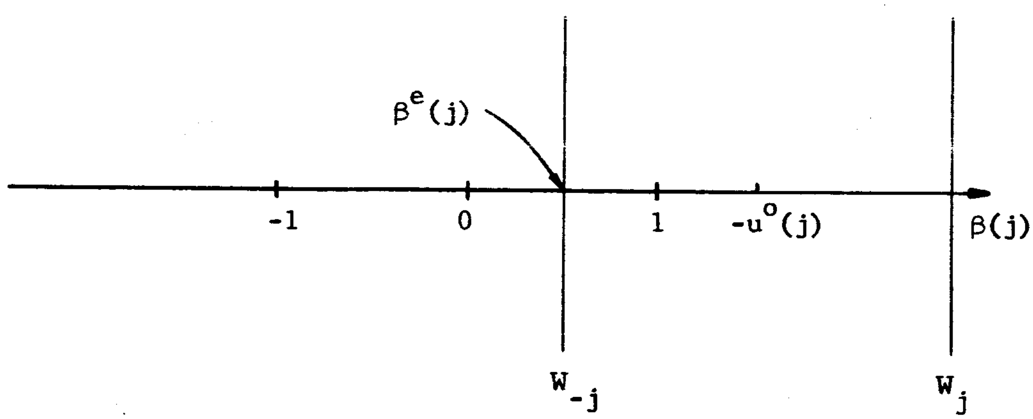
Proof. Since the origin of β-space is contained on or between the pairs of hyperplanes W_i, W_{-i}, $i = 1, 2, \ldots, N$, $i \neq j$, and is not contained on or between the pair W_j, W_{-j}, $\underline{\beta}^e$ must lie on W_j or W_{-j}. Consider Figure 39. If $u^o(j) > 1$, $\beta^e(j) = -u^o(j) + 1$, and therefore, from Equation (3-68), $u^e(j) = 1$. If $u^o(j) < -1$, $\beta^e(j) = -u^o(j) - 1$ and $u^e(j) = -1$. Therefore, $u^e(j)$ is given by Equation (3-75) and the theorem is proved.

Theorem 2 was proved by Stubberud and Swiger (36, page 405) in the solution space. As mentioned earlier, it is felt that the reasoning behind the proof can be followed more easily in β-space. Theorem 2 has immediate use. If, on following the step by step open loop control scheme demonstrated earlier, not more than one member of $\underline{u}^o$ exceeds the saturation limits at each step, Theorem 2 guarantees that the resulting $\underline{u}^e$ is optimum.

Now suppose $\left|u^o(j)\right| > 1$ for more than one integer j. Does Equation (3-75) hold for more than one value of j? Figure 38, page 102,



a. $u^o(j) > 1$



b. $u^o(j) < -1$

Figure 39. The two cases $u^o(j) > 1$, $u^o(j) < -1$ of Theorem 2.

where $u^o(3)$ and $u^o(4)$ both exceed the saturation limit, shows immediately that, in general, Equation (3-75) is not true for more than one value of j, since $|u^e(3)| < 1$. Postulate 1, based on the work of Stubberud and Swiger (36), suggests a method for finding which members of $\underline{u}^e$ are to be set equal to the saturation limit.

Let the points of tangency of the hyperellipsoid of Equation (3-51) to the hyperplane W_j and W_{-j}, j = 1, 2, ..., N, be called $\underline{\beta}_j$ and $\underline{\beta}_{-j}$ respectively. Corresponding to these (N-n) x 1 vectors $\underline{\beta}_j$ and $\underline{\beta}_{-j}$ are the N x 1 vectors $\underline{\delta}_j^+$ and $\underline{\delta}^-$, obtained via Equations (3-49) and (3-47) as

$$\underline{\delta}_j^+ = \begin{bmatrix} -H\,\underline{\beta}_j \\ \\ \underline{\beta}_j \end{bmatrix} \quad , \quad j = 1, 2, \ldots, N \quad , \tag{3-76}$$

and

$$\underline{\delta}_j^- = \begin{bmatrix} -H\,\underline{\beta}_{-j} \\ \\ \underline{\beta}_{-j} \end{bmatrix} \quad , \quad j = 1, 2, \ldots, N \quad . \tag{3-77}$$

Let the set of integers j for which $|u^o(j)| > 1$ be called J. For notational simplicity assume, without loss of generality, that $u^o(1) > 1$. Then:

<u>Postulate 1</u>. If, for all integers j in J,

$$u(1) = u^o(1) + \delta_j^+(1) \geqslant 1 \quad \text{when } u^o(j) > 1 \tag{3-78}$$

$$u(1) = u^o(1) + \delta_j^-(1) \geqslant 1 \quad \text{when } u^o(j) < -1 \tag{3-79}$$

then it follows that $u^e(1) = 1$.

The modification for the case $u^o(1) < -1$ is simply to replace $\geqslant 1$ by $\leqslant -1$ in Equations (3-78) and (3-79), and the result that follows is that $u^e(1) = -1$. Postulate 1 can be stated verbally: Suppose more than one member of $\underline{u}^o$ exceeds the saturation limit and, typically, $u^o(1) > 1$. If the additional constraint

$$u(i) = \text{sgn. } u^o(i), \quad i \text{ in } J , \tag{3-80}$$

is adjoined to Equation (2-16) in the linear energy problem, Problem (2-18), and a new $u^o(1)$ is recalculated for each separate i in J, then $u^e(1) = +1$ if each $u^o(1)$ is still greater than unity. Figure 40 gives an example, for a typical second order system with N = 4, where Postulate 1 is valid: the first member, $u^o(1)$ and the last member, $u^o(4)$, both exceed unity, and J consists of the integers 1 and 4. Then $u^e(1) = 1$ since $\underline{\beta}_4$ lies beyond W_1 giving $u(1) = u^o(1) + \delta_4^+(1) > 1$. Similarly $u^e(4) = 1$, since $\underline{\beta}_1$ lies beyond W_4, giving $u(4) = u^o(4) + \delta_1^+(4) > 1$. For a different initial state $\underline{c}$, Figure 41 gives another example where Postulate 1 is valid. Here $u^o(3) < -1$ and $u^o(4) > 1$. The set J consists of the integers 3 and 4. Then $u^e(4) = 1$ since $\underline{\beta}_{-3}$ lies beyond W_4, whereas $u^e(3) > -1$ since $\underline{\beta}_4$ lies inside W_{-3}. From these examples it might appear that Equations (3-78) and (3-79), or their equivalents when $u^o(1) < -1$, be postulated not only as sufficient conditions for $u^e(1) = 1$, but also as necessary conditions. That these are not necessary conditions may be seen by considering Figure 42, which shows $u^o(3) < -1$ and $u^o(4) > 1$. For the purposes of Postulate 1 $u(4) = u^o(4) + \delta_3^+(4) > 1$, but $u(3) = u^o(3) + \delta_4^+(3) > -1$, however, it is evident from Figure 42 that both $u^e(4) = 1$ and $u^e(3) = -1$.

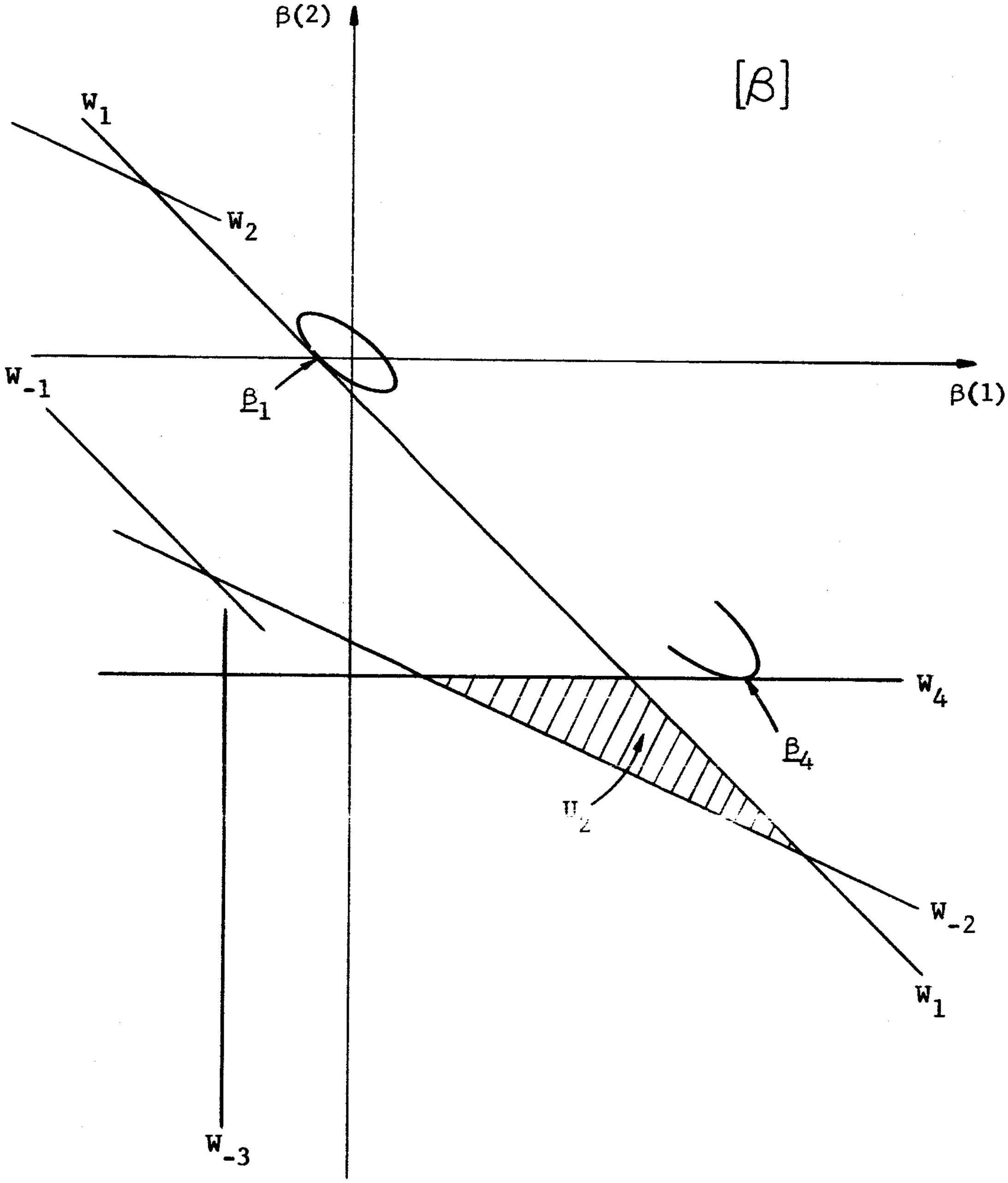


Figure 40. First example where Postulate 1 is valid.

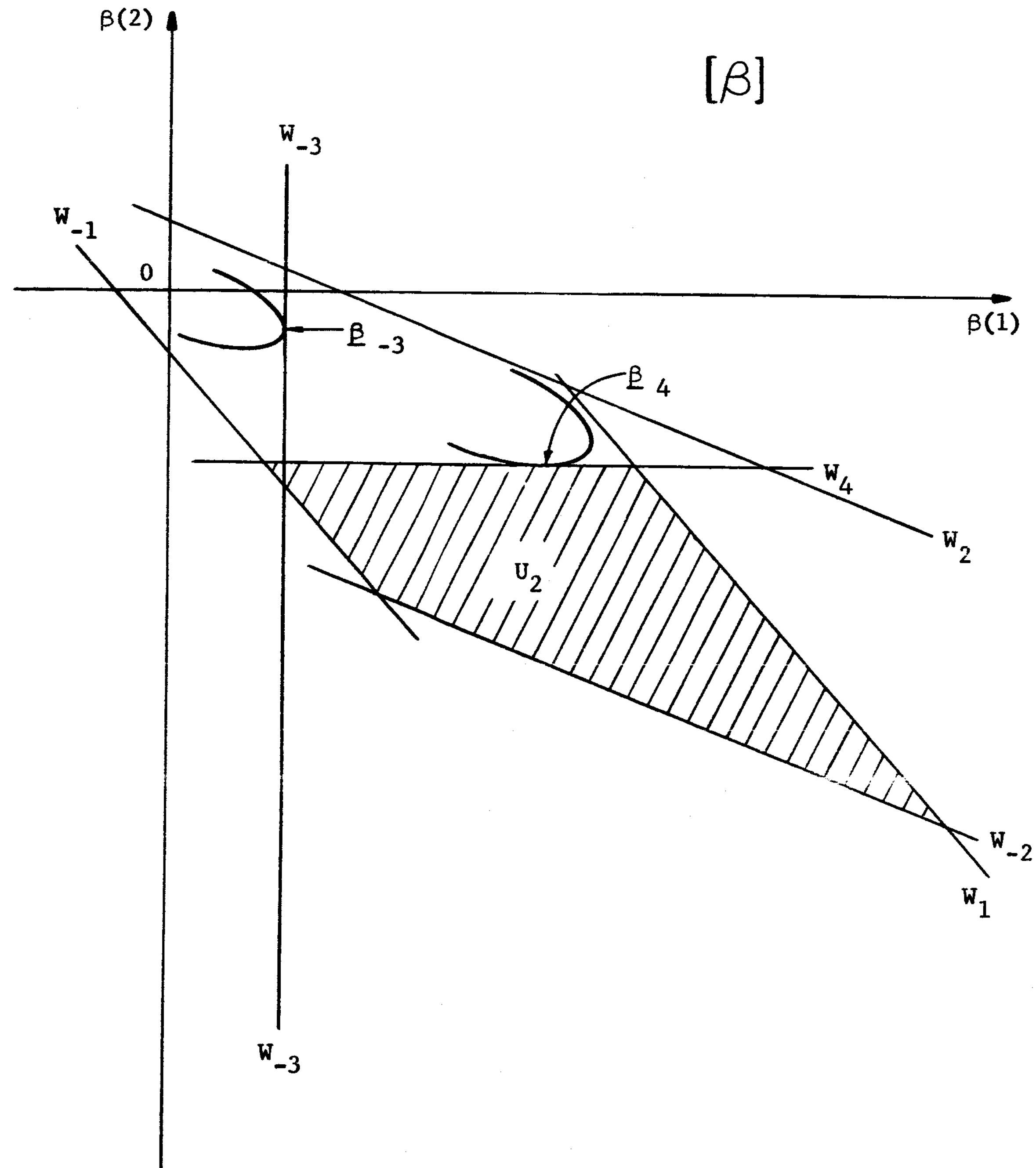


Figure 41. Second example where Postulate 1 is valid.

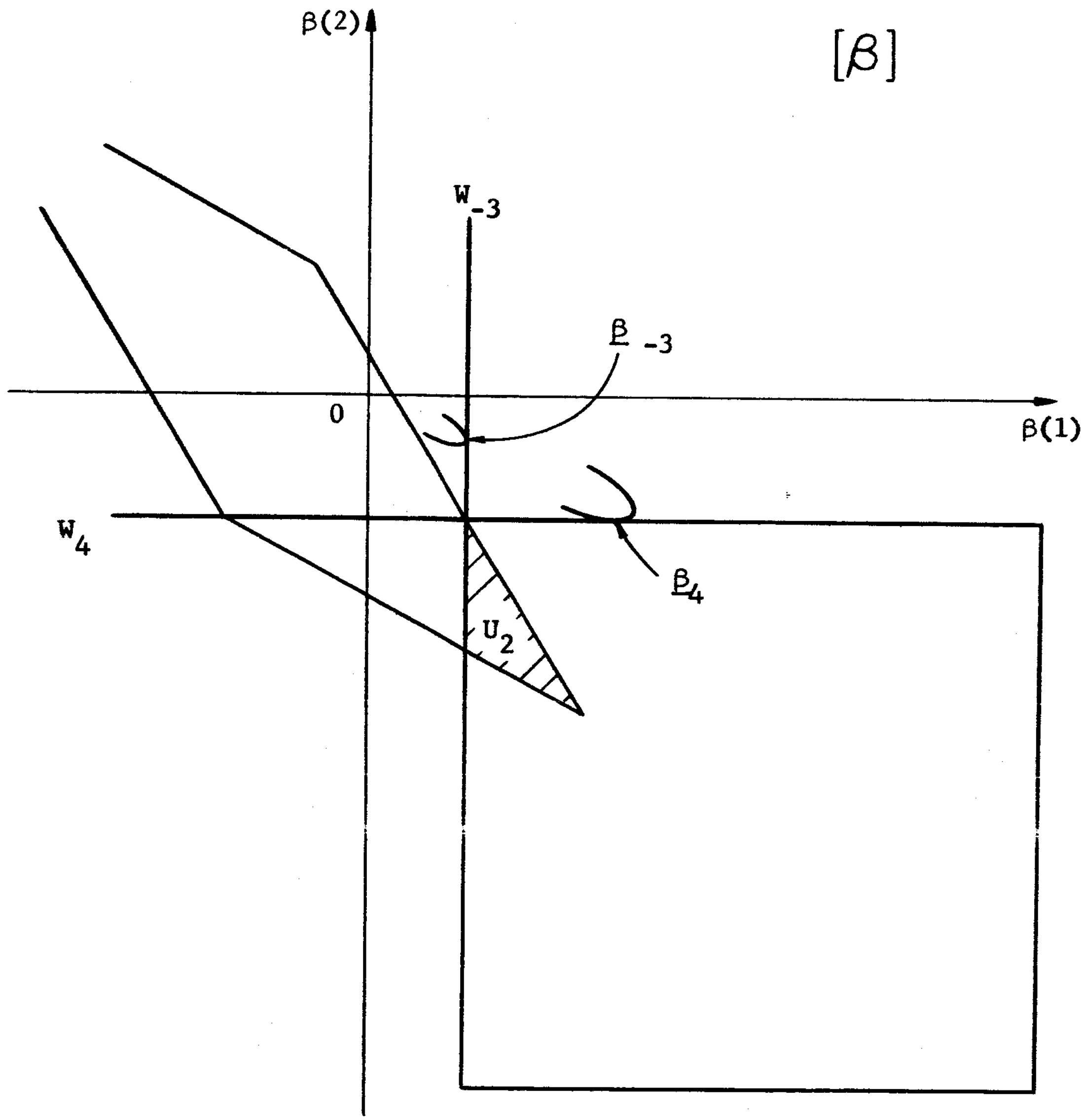


Figure 42. An example showing that the conditions of Postulate 1 are not necessary conditions.

Having given two examples where Postulate 1 is correct, it is equally possible to find examples where it is incorrect. Consider the plant given by Equation (3-59). Let $\underline{c}$ be given as

$$\underline{c} = \begin{bmatrix} -3.3 \\ -2.7125 \end{bmatrix} , \tag{3-81}$$

corresponding to an initial state on the boundary of Γ_4. The linear minimum energy input sequence is

$$\underline{u}^o = \text{col.} \left[-0.583, \quad -0.5062, \quad 1.04, \quad 1.272 \right] . \tag{3-82}$$

Figure 43 shows the corresponding sets A_2' and B_2. The set U_2 is a single point, given by

$$\underline{\beta} = \begin{bmatrix} -0.14 \\ -0.272 \end{bmatrix} . \tag{3-83}$$

Since U_2 is a single point, Equation (3-83) also gives $\underline{\beta}^e$, and therefore,

$$\underline{u}^e = \text{col.} \left[-1, \quad -1, \quad 0.9, \quad 1 \right] . \tag{3-84}$$

Postulate 1 would give the components $u^e(4) = 1$ and $u^e(3) = 1$; i.e., a $\underline{\beta}^e$ at the intersection of W_3 and W_4. This would leave the task of taking the new initial state

$$\underline{c} = \begin{bmatrix} -3.3 \\ -2.7125 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 1.9375 \end{bmatrix} + \begin{bmatrix} 2 \\ -0.25 \end{bmatrix} = \begin{bmatrix} -0.8 \\ -1.025 \end{bmatrix} , \tag{3-85}$$

into the origin to the remaining invariant vectors $\underline{h}_1$ and $\underline{h}_2$. Therefore, in this case Postulate 1 breaks down.

The set M_N includes a substantial region of states in Γ_N. Similarly, it would be valuable to know the size of the region of initial states for which Postulate 1 is valid. This extended region is

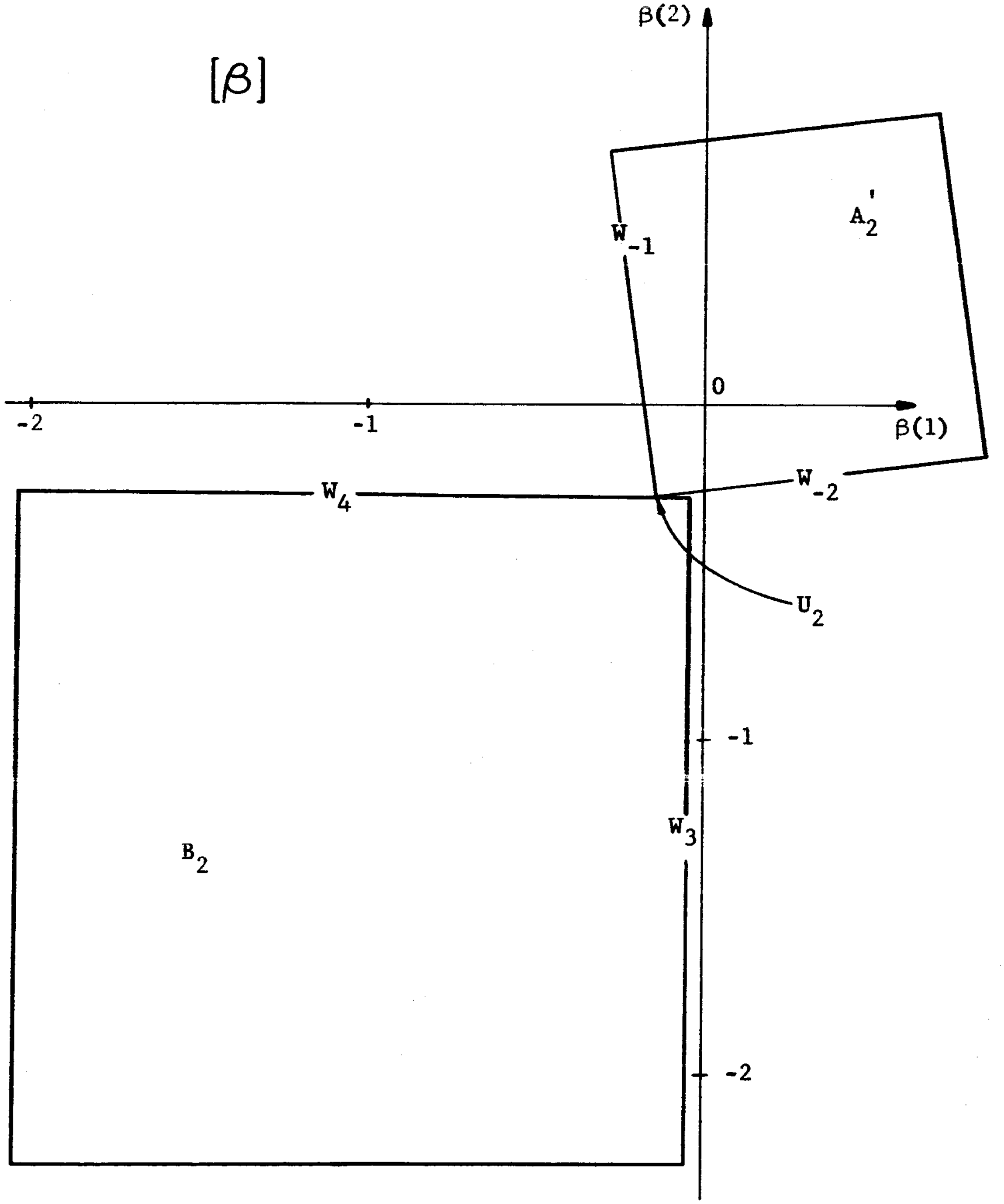


Figure 43. The set U_2 shows that Postulate 1 is not always valid.

illustrated by means of an example, which leads directly to Theorem 3. The example uses the second order plant of Equation (3-59),

$$G_p(s) = \frac{1}{(s + 3.465 + j\,14.83)(s + 3.465 - j\,14.83)} \qquad (3\text{-}86)$$

and the sampling period is again T = 1. Using the invariant vectors of Equation (3-61), the sets Γ_4 and M_4 are generated and are shown in Figure 44. The cross-hatched regions are regions where Theorem 2 applies, and therefore, where Postulate 1 is valid. The regions of initial states for which two members of $\underline{u}^o$ exceed the saturation limit are labelled A, B, A^- and B^-. Because of the symmetry, it is only necessary to consider states in A and B. The initial state of Equation (3-81) is in region A, so that, for at least one state $\underline{c}$, the postulate is invalid. The question to be discussed next is: how many other initial states in A and B have optimum input sequences whose members cannot be obtained by using Postulate 1? A short digression, resulting in a more convenient way of stating Postulate 1, is necessary before this question can be answered.

Let the point of tangency of the hyperellipsoid of Equation (3-51) with the hyperplane $\delta(j)$ = constant be called $\underline{\delta}_j$. If the hyperplane is W_j, $\underline{\delta}_j$ corresponds to $\underline{\delta}_j^+$ of Equation (3-76), and if W_{-j}, $\underline{\delta}_j$ corresponds to $\underline{\delta}_j^-$ of Equation (3-77). It is shown in Appendix B, Equation (B-46), that the i-th member of $\underline{\delta}_j$ is given by

$$\delta_j(i) = \frac{\delta_{ij} - T_{ij}}{1 - T_{jj}}\ \delta(j), \quad i,\ j = 1,\ 2,\ \ldots,\ N, \qquad (3\text{-}87)$$

where δ_{ij} is the Kronecker delta, $1 - T_{jj} > 0$ from Equation (B-53)

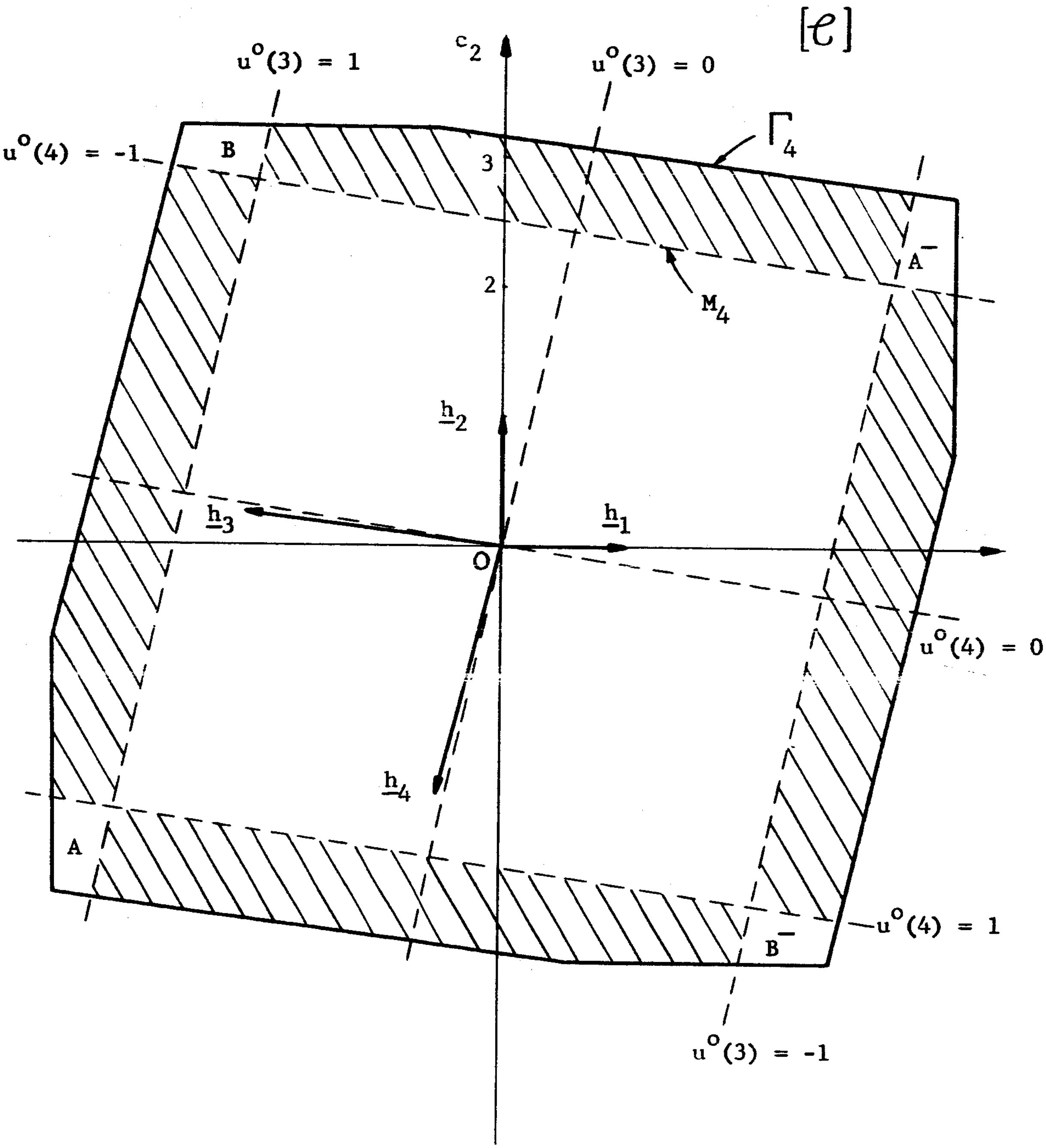


Figure 44. The regions A, B, A^- and B^- where two members of $\underline{u}^o$, for the plant of Equation (3-86), exceed the saturation limit.

and

$$T_{ij} = u^{o}(i, \underline{h}_j) \quad , \tag{3-88}$$

where the notation $u^{o}(i, \underline{h}_j)$ refers to the i-th member, i = 1, 2, ..., N, of the linear minimum energy input sequence when the initial state $\underline{c}$ is the invariant vector $\underline{h}_j$, j = 1, 2, ..., N. If the point of tangency to the hyperplane W_j is desired then

$$\delta(j) = - u^{o}(j) + 1 \quad , \tag{3-89}$$

and if the point of tangency to the line W_{-j} is desired, then

$$\delta(j) = - u^{o}(j) - 1 \ . \tag{3-90}$$

For the purposes of Postulate 1, if $u^{o}(j) > 1$ the point of tangency is to be with the hyperplane W_j, and if $u^{o}(j) < -1$ the hyperellipsoid is to be tangential to the line W_{-j} (see Equations (3-78) and (3-79)). Therefore, in either case (compare Figure 39, page 104) $\delta(j)$ in Equation (3-87) is to be given by

$$\delta(j) = \text{sgn. } u^{o}(j) - u^{o}(j) \quad . \tag{3-91}$$

Postulate 1, in its most general form, can therefore be restated as:

<u>Postulate 1a</u>. If i is in J, $u^{e}(i) = \text{sgn. } u^{o}(i)$, if, for all j in J,

$$u^{o}(i) + \delta_j(i) \quad \left.\begin{array}{ll} \geqslant 1 & \text{if } \ u^{o}(i) > 1 \\ \\ \leqslant -1 & \text{if } \ u^{o}(i) < -1 \end{array}\right\} \quad , \tag{3-92}$$

where $\delta(j)$ in Equation (3-87) is given by Equation (3-91).

Having restated Postulate 1 in the more convenient form of Postulate 1a, it is possible to continue the examination of regions A and B in Figure 44. When the initial state is in region A, $u^o(3) > 1$ and $u^o(4) > 1$. Postulate 1a states that $u^e(3) = 1$ if

$$u^o(3) - (T_{34}/1 - T_{44})\ \delta(4) \geqslant 1 \quad , \tag{3-93}$$

where $\delta(4) = 1 - u^o(4)$ is a negative quantity and $T_{34} = u^o(3, \underline{h}_4)$ is, as can be seen from Figure 44, a positive quantity. Further, from Equation (B-53), $1 - T_{jj}$ is always positive. Therefore,

$$u^o(3) - (T_{34}/1 - T_{44})\ \delta(4) > 1 \quad , \tag{3-94}$$

and $u^e(3) = 1$. Similarly, $\delta(3) = 1 - u^o(3)$ is also negative, and since $T_{43} = T_{34}$, see Equation (B-55), $u^e(4) = 1$. Thus, Postulate 1a gives, for all initial states in A,

$$u^e(3) = u^e(4) = 1 \quad . \tag{3-95}$$

Now consider initial states in region B. In this region $u^o(3) > 1$ and $u^o(4) < -1$. A straightforward calculation, using Equation (3-88), gives,

$$-T_{34}/1 - T_{44} = -0.10185 \tag{3-96}$$

$$-T_{43}/1 - T_{33} = -0.10304 \quad . \tag{3-97}$$

Postulate 1a says that $u^e(3) = 1$ if

$$u^o(3) - 0.10185\left[-1 - u^o(4)\right] \geqslant 1 \quad , \tag{3-98}$$

and $u^e(4) = -1$ if

$$u^o(4) - 0.10304\ \left[1 - u^o(3)\right] \leqslant -1 \quad . \tag{3-99}$$

Equations (3-98) and (3-99) give the three possible occurrences shown in Figure 45: when the initial state is in region B, the postulate gives either $u^e(3) < 1$, $u^e(4) = -1$, or $u^e(3) = 1$, $u^e(4) = -1$, or finally

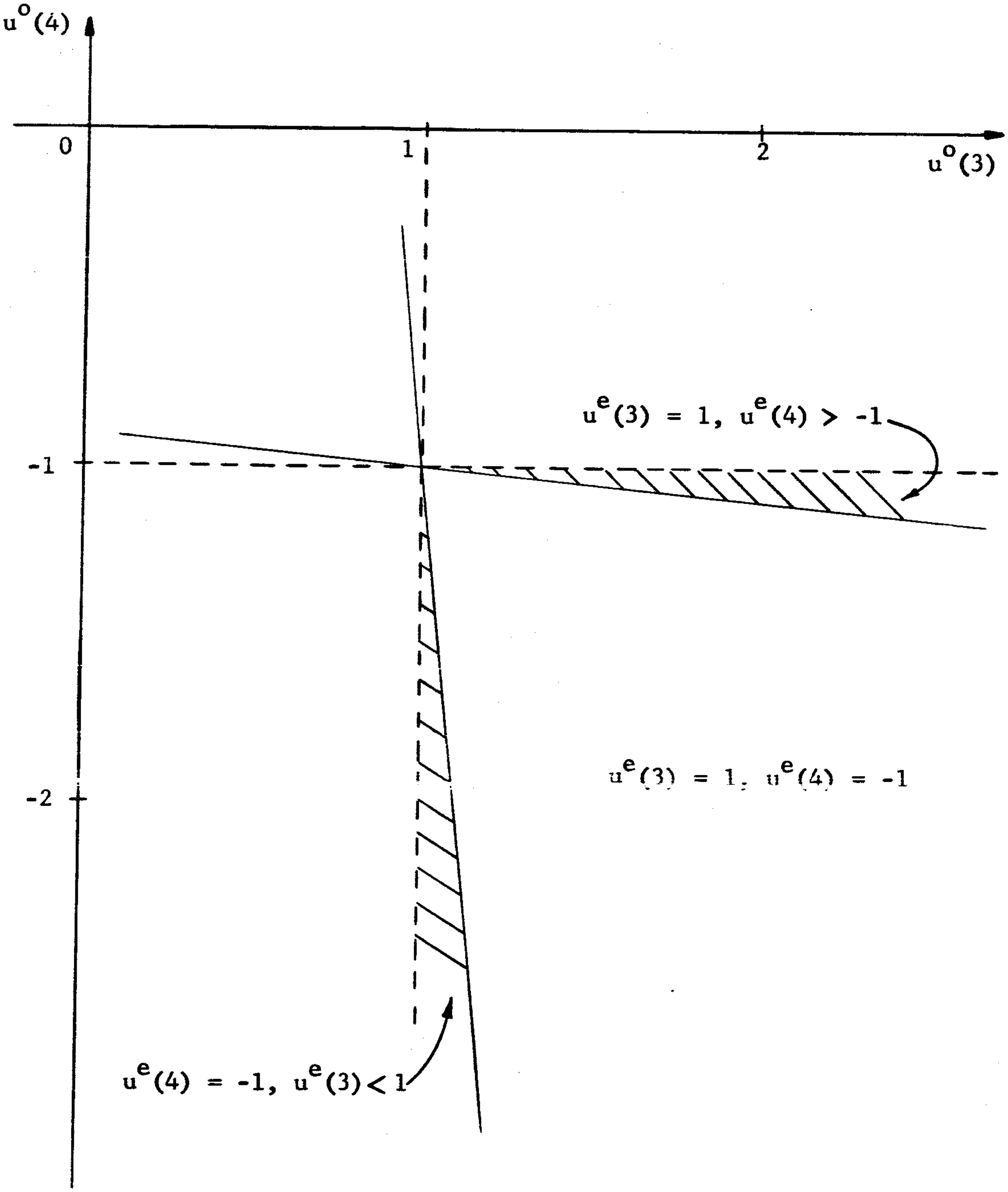


Figure 45. The possible values of $u^e(3)$ and $u^e(4)$ for initial states in region B of Figure 44, page 113, as given by Postulate 1a.

$u^e(3) = 1$, $u^e(4) > -1$. The corresponding initial states $\underline{c}$ for which these results apply are shown in Figure 46.

Regions A and B contain sub-regions of initial states for which Postulate 1a is invalid. In region A, for example, it has been shown, in Equation (3-95), that the postulate requires $u^e(3) = 1$. Setting $u^e(3) = 1$ leaves the invariant vectors $\underline{h}_1$, $\underline{h}_2$ and $\underline{h}_4$ to represent the new initial state,

$$\underline{c}' = \underline{c} - \underline{h}_3 \quad . \tag{3-100}$$

Therefore, if $\underline{c}'$ does not lie in the set Γ_3', formed from $\underline{h}_1$, $\underline{h}_2$ and $\underline{h}_3$ as, see Figure 47,

$$\Gamma_3' = \left(\underline{c}' \,\middle|\, \underline{c}' = u(1)\underline{h}_1 + u(2)\underline{h}_2 + u(4)\underline{h}_4 ; \ \left|u(j)\right| \leqslant 1, \ j = 1,2,4 \right) , \tag{3-101}$$

then Postulate 1a is invalid, since it is impossible to take $\underline{c}'$ into the origin with any input sequence satisfying the saturation constraint.

For an initial state $\underline{c}$ in region B, setting $u^e(4) = -1$ gives

$$\underline{c}' = \underline{c} + \underline{h}_4 \quad . \tag{3-102}$$

If $\underline{c}'$ does not lie in the set Γ_3,

$$\Gamma_3 = \left(\underline{c}' \,\middle|\, \underline{c}' = u(1)\underline{h}_1 + u(2)\underline{h}_2 + u(3)\underline{h}_3 ; \ \left|u(j)\right| \leqslant 1, \ j = 1,2,3 \right) , \tag{3-103}$$

see Figure 47, the postulate is also invalid, for the same reason.

Figure 47 also shows, by the cross-hatched areas, the subregions of A, B, A^- and B^- for which Postulate 1a is invalid. In this example, these regions are very small when compared with the size of Γ_4.

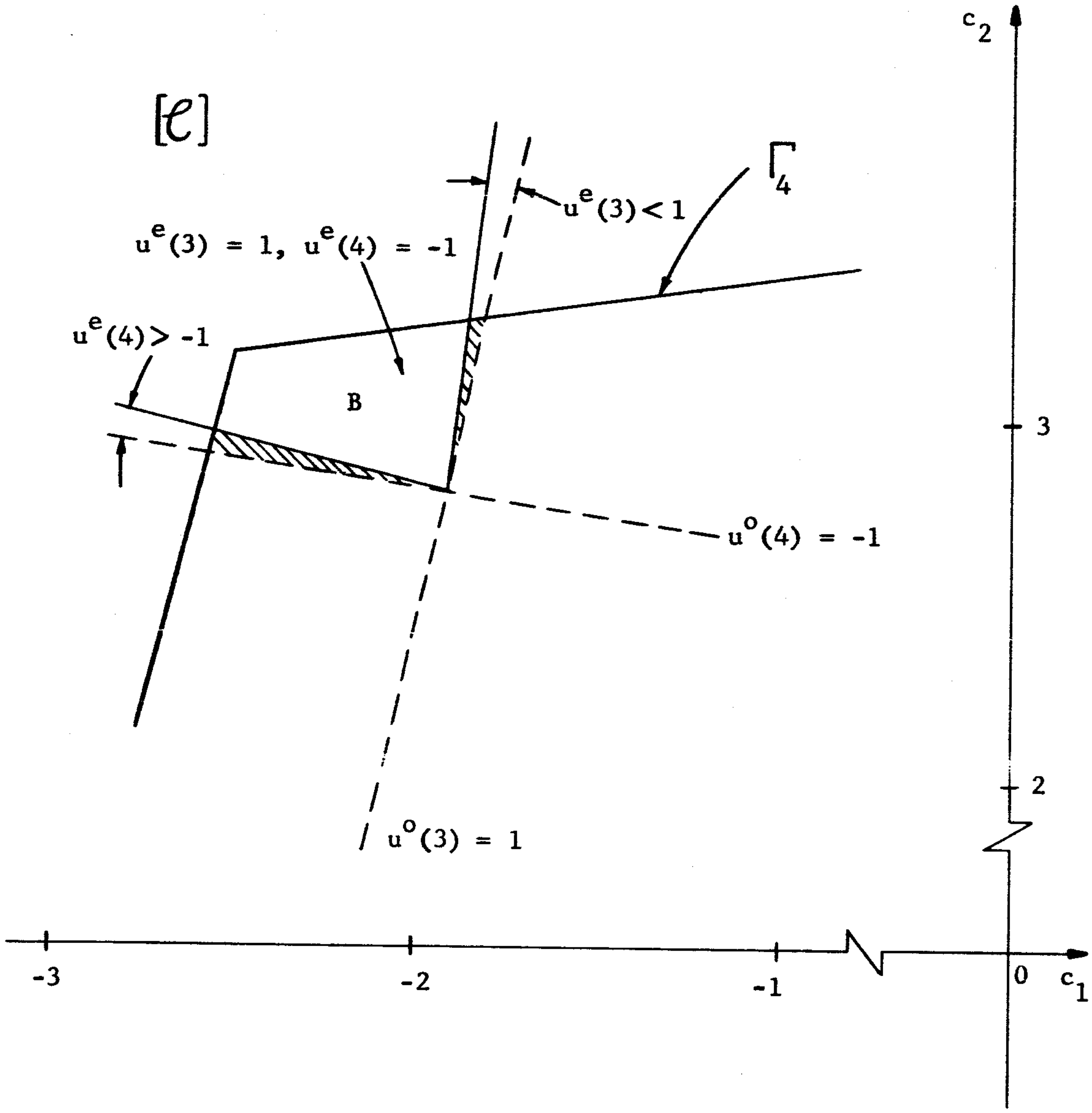


Figure 46. The possible values of $u^e(3)$ and $u^e(4)$ for initial states in region B, as given by Postulate 1a.

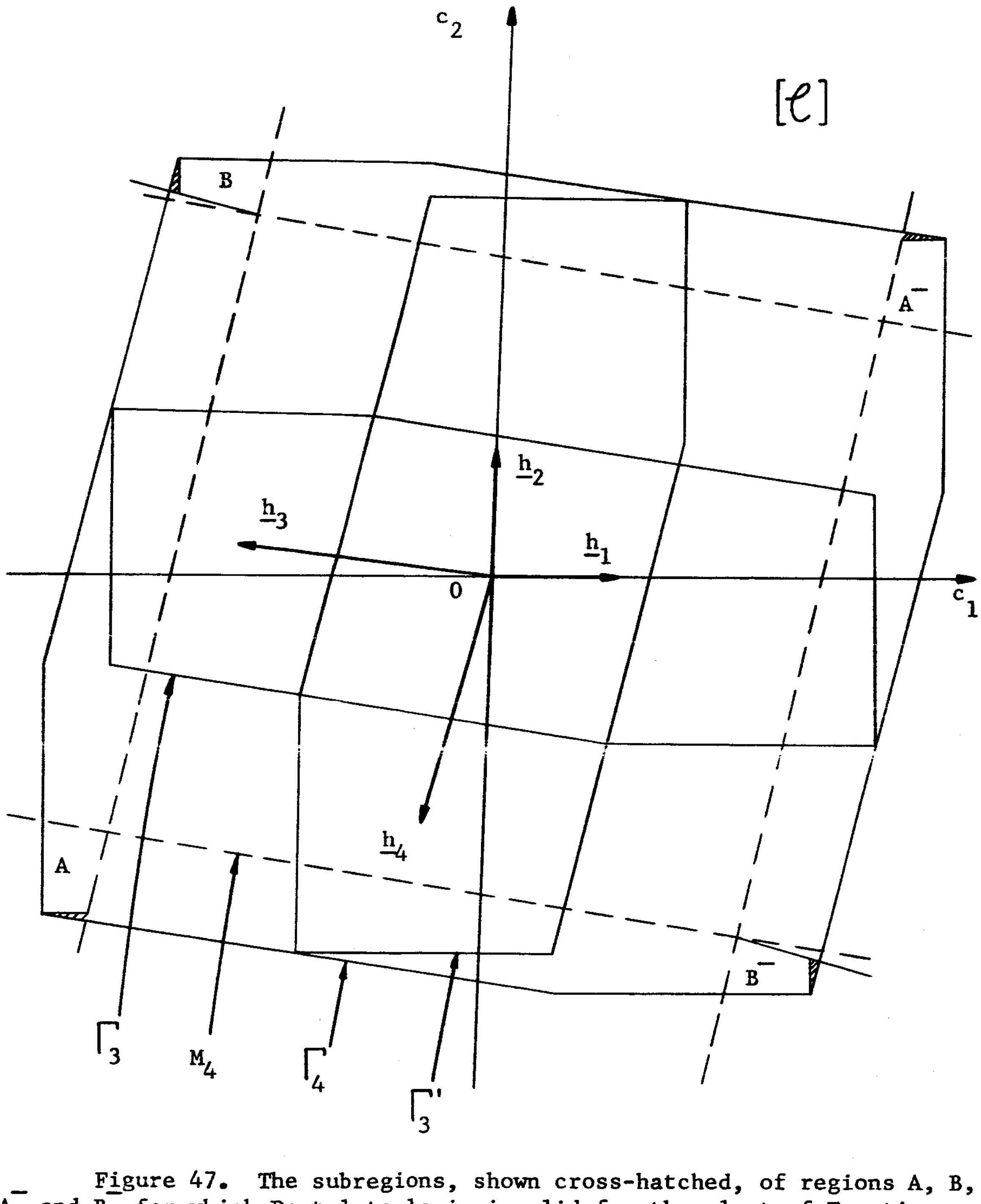


Figure 47. The subregions, shown cross-hatched, of regions A, B, A^- and B^- for which Postulate 1a is invalid for the plant of Equation (3-86) with N = 4.

In general, the following theorem, proved in Appendix B, can be of significant help in obtaining the sequence $\underline{u}^e$.

Theorem 3. Given an initial state $\underline{c}$ in Γ_N but not in M_N. For a given i in the set J, calculate, from Equation (3-87), $\delta_j(i)$ for all j in J. Then

$$u^e(i) = \text{sgn. } u^o(i) \tag{3-104}$$

if, for all integers j in J,

$$u^o(i) + \delta_j(i) \quad \left.\begin{array}{l} \geqslant 1 \quad \text{if} \quad u^o(i) > 1 \\ \\ \leqslant -1 \quad \text{if} \quad u^o(i) < -1 \end{array}\right\} , \tag{3-105}$$

where $\delta_j(i)$ is obtained from Equation (3-87) with $\delta(j)$ given by Equation (3-91), and

$$\underline{c}' = \underline{c} - \left[\text{sgn. } u^o(i)\right] \underline{h}_i \tag{3-106}$$

is in the set Γ'_{N-1}, where

$$\Gamma'_{N-1} = \left(\underline{c}' \,\middle|\, \underline{c}' = \sum_{\substack{j=1 \\ j\neq i}}^{N} u(j)\underline{h}_j ; \quad |u(j)| \leqslant 1, \ j = 1, 2, \ldots, N \right) . \tag{3-107}$$

This theorem is useful if it can be shown that $\underline{c}'$ is in Γ'_{N-1}. The only general way to find out if $\underline{c}'$ is in Γ'_{N-1} is to actually go through the step by step open loop control procedure discussed earlier. If a sequence results which takes the initial state $\underline{c}$ to the origin, it is an optimum sequence. If, however, N having been reduced to n or less, it is found that $\underline{c}$ cannot be represented with the remaining invariant vectors then one or more members of the input sequence must have been erroneously set equal to the saturation limit. There are several cases, however, where Theorem 3 can be used to generate the optimum input sequence.

First Order Systems

It can now be demonstrated that Equation (3-10) may be used to generate an optimum sequence for first order systems. Without loss of generality, consider stable first order systems with initial states $\underline{c} > 0$. It was shown, Equation (3-4), that if $\underline{c}$ is not in M_N, at least the last member of $\underline{u}^o$ exceeds the saturation limit, $|u^o(N)| > 1$. Then $\underline{c}' = \underline{c} - \underline{h}_N$ is certainly in Γ_{N-1} if $\underline{c}$ is in Γ_N. Furthermore, all the members of the input sequence $\underline{u}^o$ are positive and $u^o(i, \underline{h}_j)$ is always positive. Therefore, Equations (3-105) and (3-107) are satisfied, and Equation (3-10) results.

Second Order Systems

The only results of significant generality which have been obtained for second order systems are for plants with real poles, at least one of which corresponds to an integration, and for plants with complex poles which have been "tuned" (47, page 95).

<u>Second order plants with integration</u>. Consider the stable second order plant,

$$G_p(s) = \frac{1}{s(s+\lambda)} \quad , \quad \lambda \geqslant 0 \; , \tag{3-108}$$

with invariant vectors, see Table I, Appendix A, page 252, given by

$$\underline{h}_{2+j} = \begin{bmatrix} -S(j) \\ \\ S(j)+1 \end{bmatrix} \quad , \quad j = 1, 2, 3, \ldots, \tag{3-109}$$

where

$$S(j) = \frac{e^{\lambda T}(e^{j\lambda T} - 1)}{e^{\lambda T} - 1} \; . \tag{3-110}$$

For this plant, the set M_N is bounded by the lines

$$u^o(1) = \pm 1, \quad u^o(N) = \pm 1, \tag{3-111}$$

which can best be seen by considering the lines $\underline{h}_j^t \, \underline{a}^o = 1$, $j = 1, 2, 3, \ldots$, shown in Figure 48. The lines $\underline{h}_N^t \, \underline{a}^o = \pm 1$ and $\underline{h}_1^t \, \underline{a}^o = \pm 1$ form the boundary of the set L_N in $\mathcal{L}^o$-space, and therefore, the lines of Equation (3-111) are the boundary lines of M_N in $\mathcal{C}$-space.

Assuming that $\underline{c}$ is in Γ_N but not in M_N, one or more of the members of $\underline{u}^o$ will exceed the saturation limit. It will be shown that, for a second order system with integration,

$$\text{if} \quad \left|u^o(j)\right| > 1 \; , \quad u^e(j) = \operatorname{sgn.} u^o(j) \; ; \tag{3-112}$$

i.e., that Equation (3-10) is applicable to such systems. It must, therefore, be shown that any initial state $\underline{c}$ in Γ_N, but not in M_N, satisfies Equations (3-105) and (3-107). Consider the lines $u^o(j) = 1$, $j = 1, 2, \ldots, N$. All these lines pass through the point on the boundary of Γ_N given by

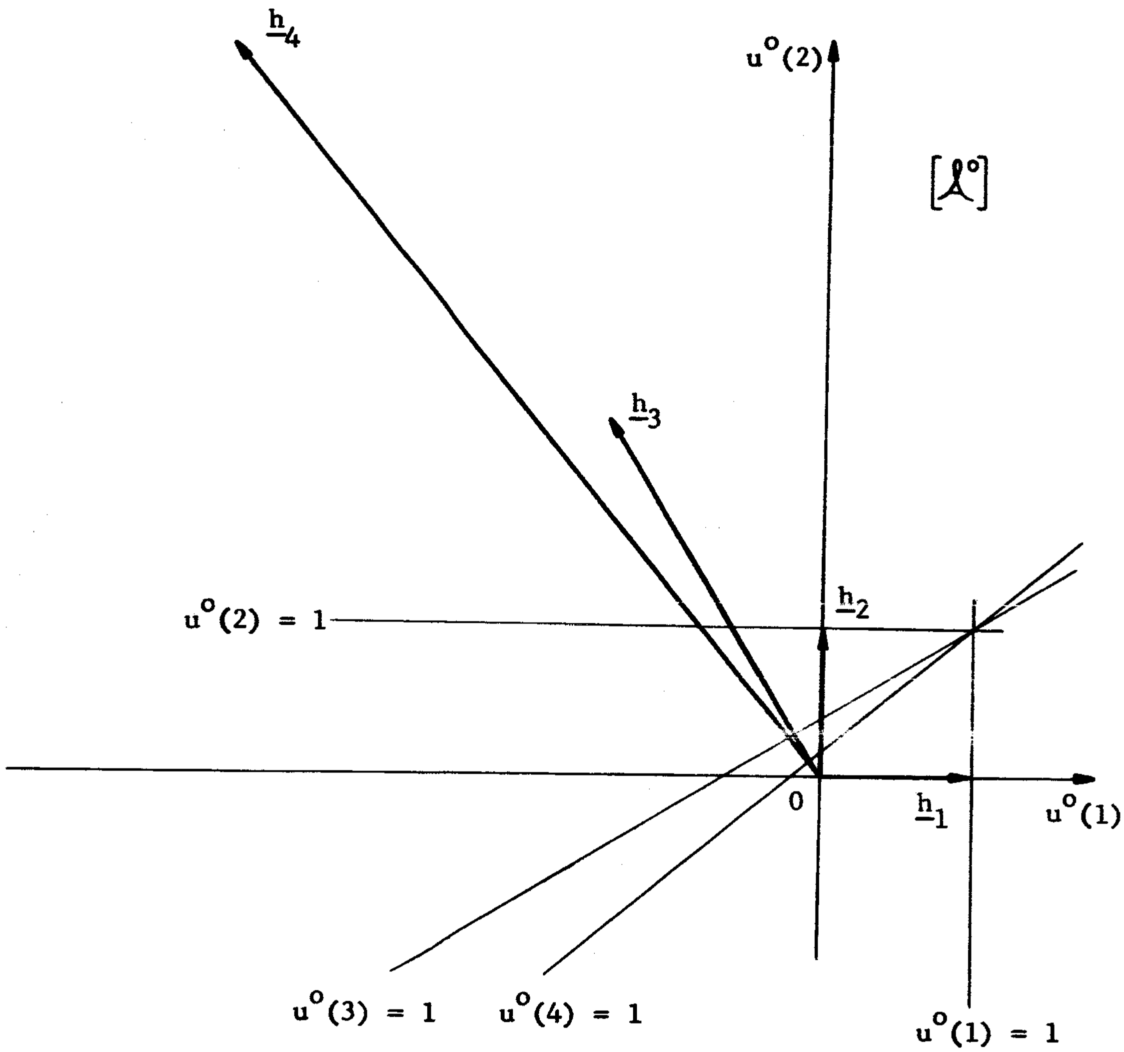


Figure 48. The lines $\underline{h}_j^t\ \underline{a}^o = u^o(j) = 1$ for the plant $1/s(s + \lambda)$.

$$\underline{c} = \sum_{j=1}^{N} \underline{h}_j \quad , \tag{3-113}$$

and by symmetry the lines $u^o(j) = -1$, $j = 1, 2, \ldots, N$, pass through the opposite corner of Γ_N given by

$$\underline{c} = - \sum_{j=1}^{N} \underline{h}_j \quad . \tag{3-114}$$

This can readily be seen by considering the set L_N: for example, in Figure 47, page 119, the lines $\underline{h}_j^t\,\underline{a}^o = 1$ all pass through the point $a_1^o = 1$, $a_2^o = 1$. The lines of Equation (3-111) partition Γ_N into six regions of interest, A, B, C, A^-, B^- and C^-. By symmetry, only the regions A, B and C need be considered. These regions are given by initial states lying in Γ_N and having linear optimum input sequences, $\underline{u}^o$, that satisfy respectively,

$$u^o(1) \geqslant 1, \quad u^o(N) \leqslant -1, \tag{3-115}$$

$$u^o(N) \geqslant 1, \quad u^o(1) \geqslant -1, \tag{3-116}$$

$$u^o(N) \geqslant 1, \quad u^o(1) \leqslant -1. \tag{3-117}$$

Figure 49 shows these regions for the plant

$$G_p(s) = \frac{1}{s^2} \quad , \tag{3-118}$$

with N = 4.

It is a straightforward, but tedious, matter to show that in general B contains the lines $u^o(j) = 1$, $j = N, N-1, \ldots, k$, when

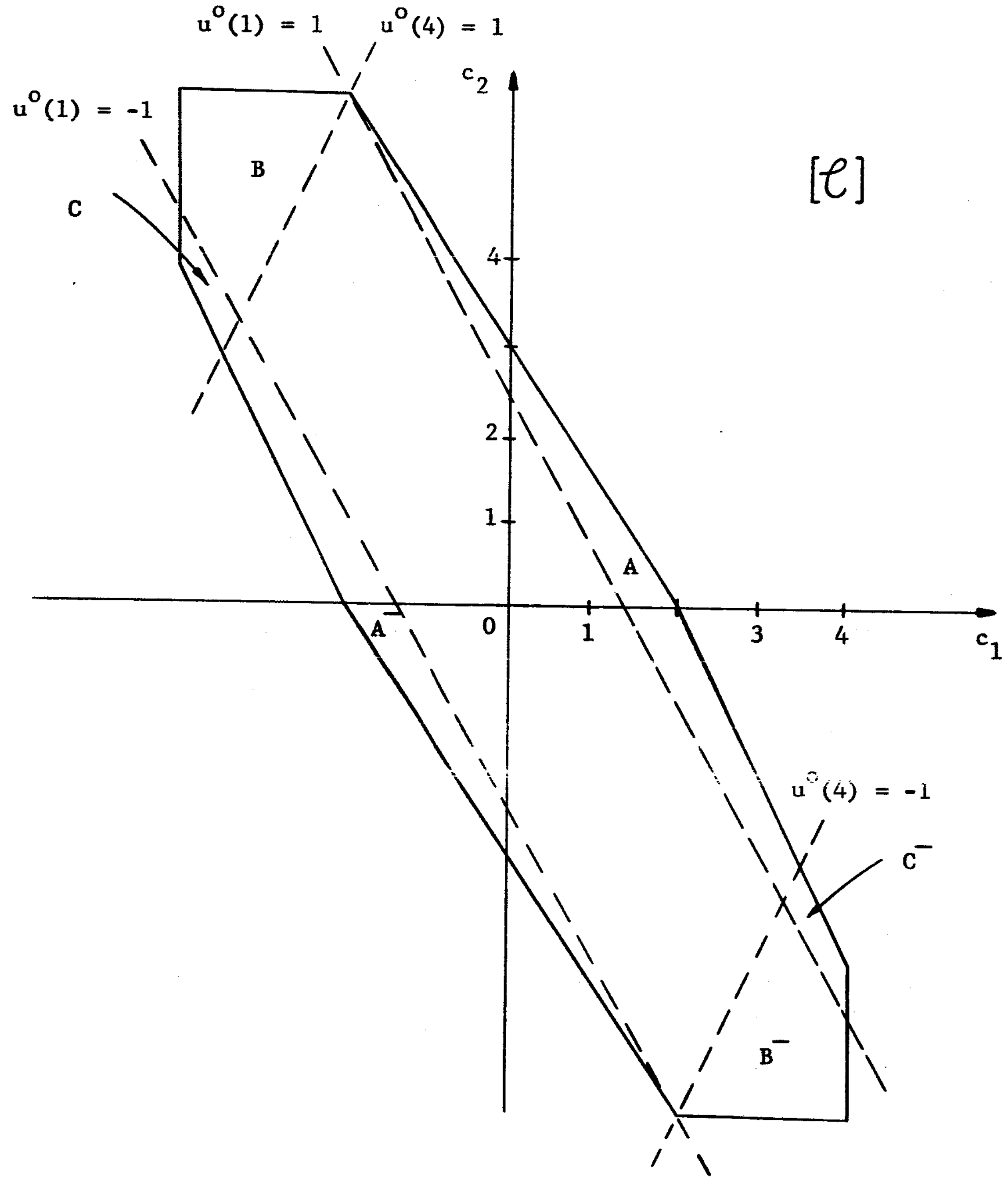


Figure 49. The regions A, B and C and their symmetrical counterparts for the plant $1/s^2$.

$$S(k)\left[1+\sum_{i=1}^{N-2}\left[S(i)+1\right]\right] \geqslant \sum_{i=1}^{N-2} S(i)\left[S(i)+1\right], \qquad (3\text{-}119)$$

and that A contains the lines $u^o(j) = 1$, $j = 1, 2, \ldots, m$, when

$$\frac{\sum_{i=1}^{N-2} S(i)\left[S(i)+1\right] - S(m)\left[1+\sum_{i=1}^{N-2}\left[S(i)+1\right]\right]}{1+\sum_{i=1}^{N-2} S^2(i) + S(m)\left[1-\sum_{i=1}^{N-2} S(j)\right]} \geqslant \frac{S(N-2)+1}{S(N-2)} .$$

(3-120)

In Equations (3-119) and (3-120), the equality holds when either $u^o(k) = 1$ or $u^o(m) = 1$ lie on the boundary of Γ_N. Figures 50 and 51 show how these equations have been used to calculate, for different values of N, which members of $\underline{u}^o$ saturate when $\underline{c}$ is in A or B. Figure 50 corresponds to the plant

$$G_p(s) = \frac{1}{s^2} , \qquad (3\text{-}121)$$

and Figure 51 to the plant

$$G_p(s) = \frac{1}{s(s+\lambda)} , \qquad e^{\lambda T} = 2 . \qquad (3\text{-}122)$$

It can be shown that if $u^o(j)$ lies in A, $j = 1, 2, \ldots, m$, $u^o(i, \underline{h}_j) > 0$ for i,j in 1, 2, ..., m. Similarly if $u^o(j) = 1$, $j = N, N-1, \ldots, k$, lies in B, $u^o(i, \underline{h}_j) > 0$ for i,j in N, N-1, ..., k. Further, $\delta(j)$ as given by Equation (3-91) is always negative for $u^o(j) = 1$ in A or B. Therefore, if $\underline{c}$ lies in A or B,

$$u^o(i) + \delta_j(i) \geqslant 1 \qquad (3\text{-}123)$$

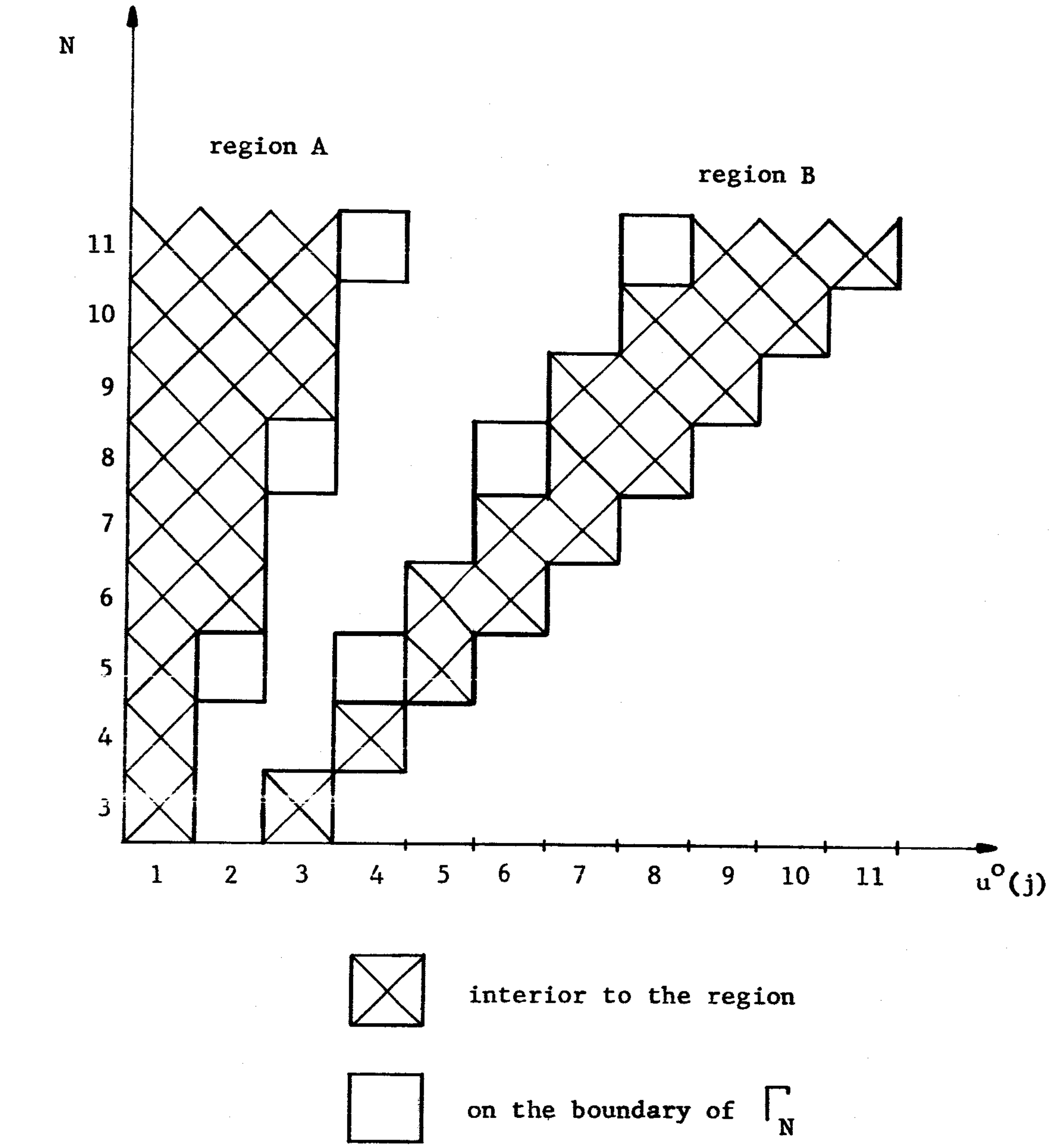


Figure 50. The lines $u^o(j) = 1$ falling in regions A and B as a function of N, for the plant of Equation (3-121).

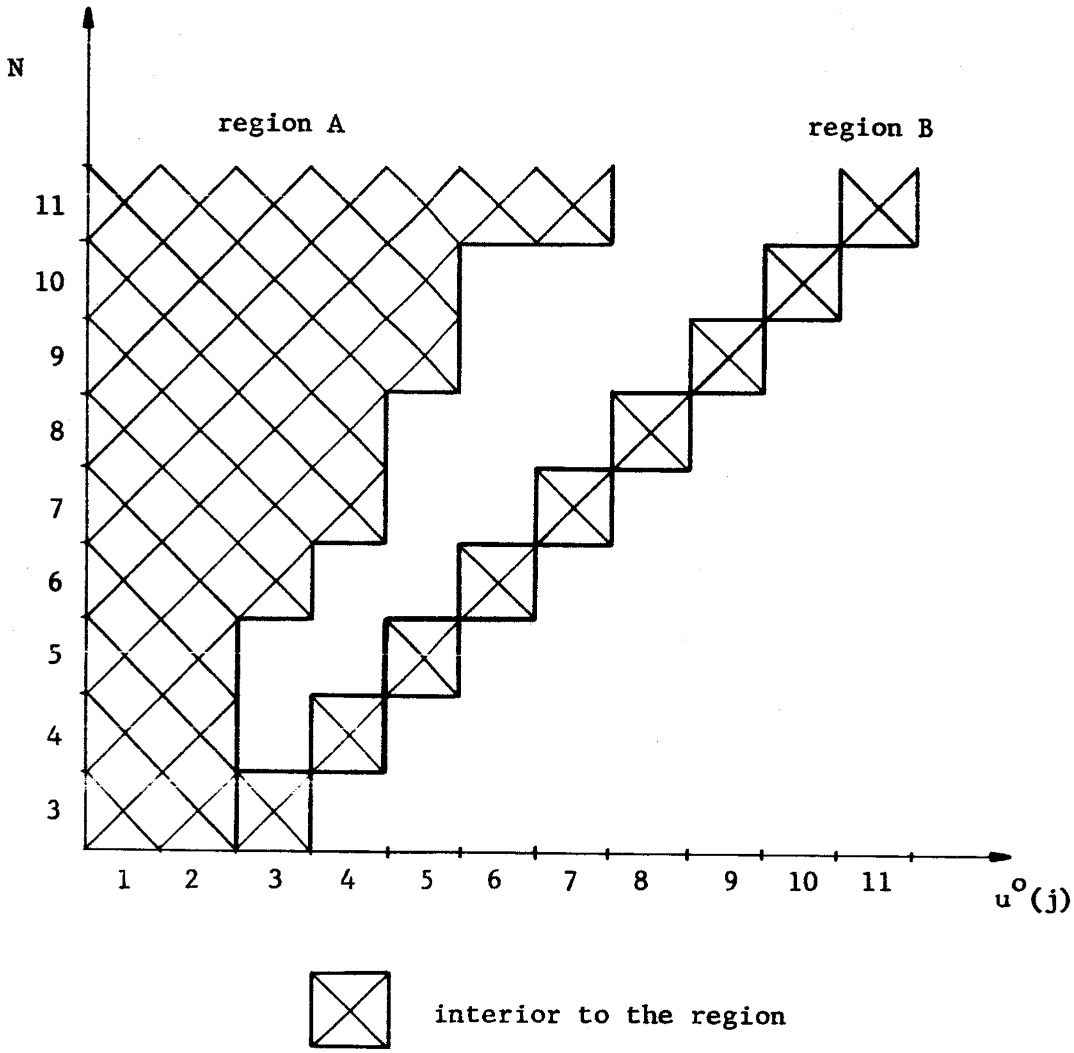


Figure 51. The lines $u^o(j) = 1$ falling in regions A and B, as a function of N, for the plant of Equation (3-122).

for any members of $\underline{u}^o$ that exceed the saturation limit. It can also be shown that Equation (3-107) is satisfied. The conditions of Theorem 3 are, therefore, satisfied for any initial state in A or B. In region C, $\delta(j)$ is positive if $u^o(j) < -1$, $j = 1, 2, \ldots$, and is negative if $u^o(j) > 1$, $j = N, N-1, \ldots$. It can be shown that $u^o(i, \underline{h}_j)$ is negative if $\underline{c}$ is in region C and $u^o(i) < -1$, $u^o(j) > 1$. Therefore,

$$u^o(i) + \delta_j(i) < -1 \quad \text{if} \quad u^o(i) < -1 \quad , \tag{3-124}$$

$$u^o(j) + \delta_i(j) > 1 \quad \text{if} \quad u^o(j) > 1 \quad . \tag{3-125}$$

It can further be shown that Equation (3-107) is also satisfied if $\underline{c}$ is in region C. It may, therefore, be concluded that, if $\underline{c}$ is in A, B or C (A^-, B^- and C^-),

$$\text{if, for any } i, \quad |u^o(i)| > 1, \quad u^e(i) = \text{sgn. } u^o(i) \quad . \tag{3-126}$$

This result enables the optimum input sequence to be obtained by a step by step procedure. For example, suppose, for the plant of Equation (3-121), that $\underline{c}$ is in B and N is given as N = 10. It is possible, see Figure 50, that $u^o(10) > 1$, $u^o(9) > 1$ and $u^o(8) > 1$. Suppose all three do saturate. Equation (3-125) guarantees that $u^e(10) = u^e(9) = u^e(8) = 1$. Since $\underline{c}' = \underline{c} - \underline{h}_{10} - \underline{h}_9 - \underline{h}_8$ lies in Γ_7, the sequence $\underline{u}^o$ for the new initial state $\underline{c}'$ with N = 7 may have $u^o(7) > 1$ and $u^o(6) > 1$. Suppose that this happens and gives $u^o(7) = u^o(6) = 1$. The state $\underline{c}' - \underline{h}_6 - \underline{h}_7$ lies in Γ_5. Figure 50 shows that the input $u^o(5)$ may saturate, but that $u^o(4) \leqslant 1$. Suppose $u^e(5) = 1$. The state $\underline{c}' - \underline{h}_6 - \underline{h}_7 - \underline{h}_5$ lies in Γ_4 where $u^o(4)$ may saturate. To terminate the procedure, suppose that $u^o(4) \leqslant 1$. Then the problem is solved, since this latest state must lie in M_4. A method of closed loop control for plants with integration is considered in Chapter V.

Second order plants with tuned complex poles. Consider the plant

$$G_p(s) = \frac{1}{(s + a + jb)(s + a - jb)} , \tag{3-127}$$

with invariant vectors given by,

$$\underline{h}_{n+j} = \begin{bmatrix} -e^{(j+1)aT} \dfrac{\sin jbT}{\sin bT} \\ \\ e^{jaT} \dfrac{\sin(j + 1)bT}{\sin bT} \end{bmatrix} , \quad j = 1, 2, \ldots . \tag{3-128}$$

Nelson (47, page 95) observed that if bT is adjusted so that

$$bT = \frac{m\pi}{2} , \quad m = 1, 2, \ldots, \tag{3-129}$$

the canonical vectors (and, therefore, the invariant vectors) become mutually orthogonal. When the system satisfies Equation (3-129), Nelson referred to the plant as being "tuned". Let

$$bT = \frac{\pi}{2} , \tag{3-130}$$

which may be accomplished by adjustment of either the sampling period or b. Minor loop feedback might be used to modify b. With a tuned plant, Equation (3-130) being satisfied, the invariant vectors become

$$\underline{h}_3 = \begin{bmatrix} -e^{2aT} \\ \\ 0 \end{bmatrix} , \quad \underline{h}_4 = \begin{bmatrix} 0 \\ \\ -e^{2aT} \end{bmatrix} , \quad \underline{h}_5 = \begin{bmatrix} e^{4aT} \\ \\ 0 \end{bmatrix} , \ldots . \tag{3-131}$$

Figure 52 shows the invariant vectors for a stable plant with $a > 0$. If

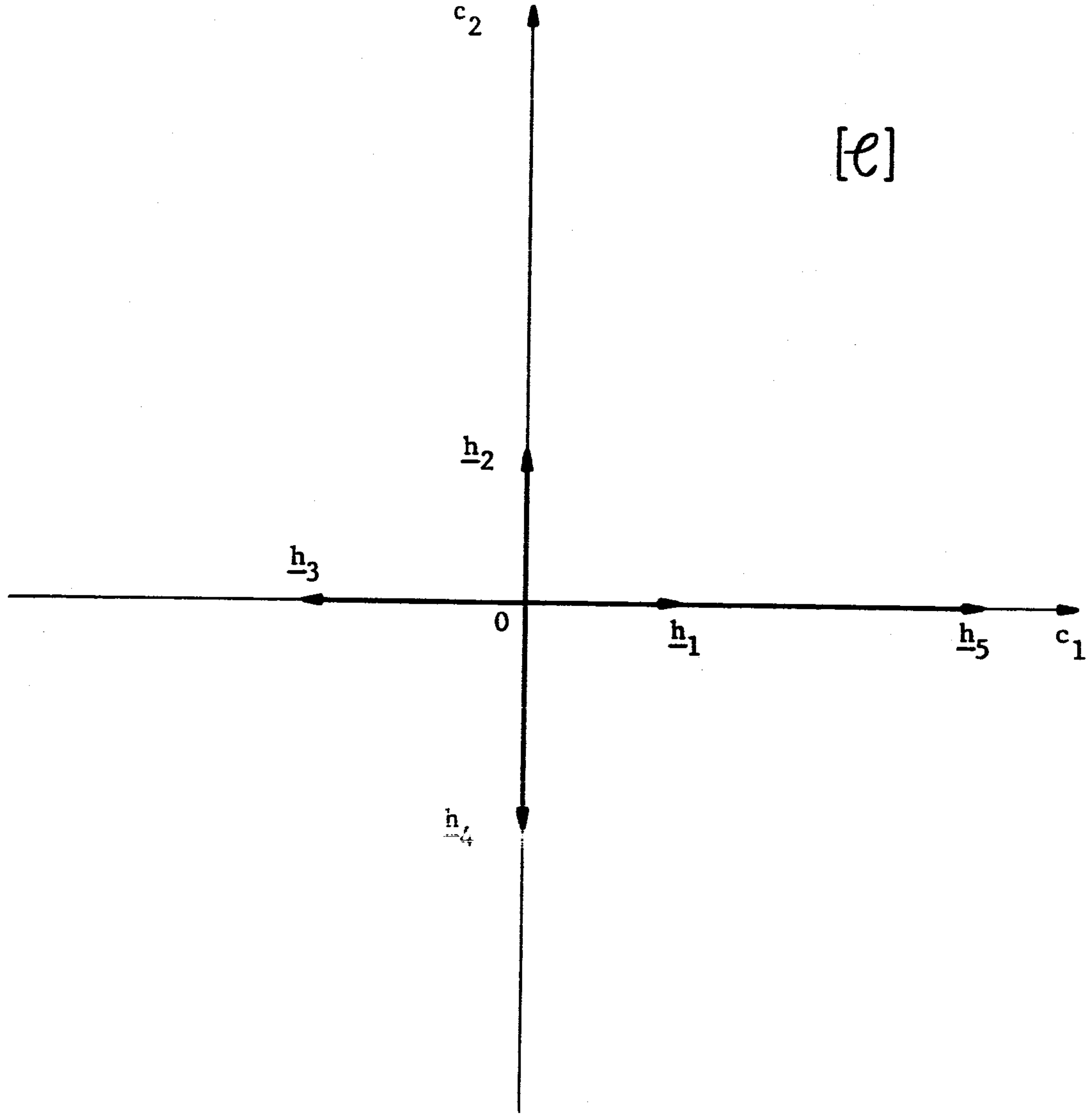


Figure 52. The invariant vectors for a stable underdamped second order plant with tuning.

$a = 0$ the tuned system has all its invariant vectors of unit length, a particularly simple configuration.

By using a method analogous to that used for plants with integration, it can be shown that Equation (3-126) is also valid for tuned plants. Even though the orthogonality of the invariant vectors makes this a simpler task than before, such considerations are not necessary. Consider the initial state $\underline{c}$ represented by the N invariant vectors $\underline{h}_1$, $\underline{h}_2, \ldots, \underline{h}_N$:

$$\underline{c} = \begin{bmatrix} c_1 \\ \\ c_2 \end{bmatrix} = \sum_{j=1}^{N} u(j)\, \underline{h}_j \quad . \tag{3-132}$$

This representation can be split into two parts,

$$c_1 = \sum_{j=1}^{N} u(j)\, \underline{h}_j \quad , \quad j \text{ odd} \quad , \tag{3-133}$$

$$c_2 = \sum_{j=2}^{N} u(j)\, \underline{h}_j \quad , \quad j \text{ even} \quad . \tag{3-134}$$

From Equations (3-128) and (3-130), compare Figure 52,

$$c_1 = u(1) + \sum_{j=3}^{N} u(j)(-1)^{(j-1)/2}\, e^{(j-1)aT} \quad , \quad j \text{ odd} \; , \tag{3-135}$$

$$c_2 = u(2) + \sum_{j=4}^{N} u(j)(-1)^{(j-2)/2}\, e^{(j-2)aT} \quad , \quad j \text{ even} \; . \tag{3-136}$$

Equations (3-135) and (3-136) are very similar to two separate first order systems, one with an initial state c_1 and the other with an initial state c_2. Although there are no real first order plants which could fit the form of Equations (3-135) and (3-136), it is clear from either the intuitive arguments or Theorem 3, that Equation (3-136) is valid for these two artifical first order systems and, therefore, for the entire tuned system. Therefore, the step by step procedure can be applied to generate the optimum sequence, $\underline{u}^e$, in an open loop manner. Closed loop control is considered for tuned systems in Chapter V.

General Second and Higher Order Systems

Second order plants with integration and underdamped plants with tuning by no means exhaust the class of second order systems. Plants with two real non-zero poles or plants with untuned complex poles are quite common. To date, there is no general way of guaranteeing that Equation (3-126), or even the more comphrehensive Postulate 1a, will generate the optimal sequence $\underline{u}^e$. However, for such second order systems, and the general n-th order system, if, on using Postulate 1a repeatedly, the initial state can be brought into the origin, then the input sequence so generated is optimum. Furthermore, since, in the example at least, the regions where Postulate 1a is not valid are so small, see Figure 47, page 119, it does seem reasonable to offer Postulate 1a as having a high probability of success. If Postulate 1a is not valid for the particular initial state, the only reasonable recourse is to use general nonlinear programming methods.

CHAPTER IV

THE MINIMUM FUEL PROBLEM WITH INPUT SATURATION

I. INTRODUCTION

The minimum fuel problem with input saturation is considered initially for first order systems. The set F_N, analogous to the set M_N of the minimum energy problem, is introduced for these first order systems. The telescoping rod analogy is then used to obtain the fuel optimum input sequence. The problems of higher order systems can be envisioned in β-space, but are more conveniently considered in the partitioned solution space. Results of significant generality have been obtained only for second order systems, where the fuel optimum sequence is considered by closed loop methods. The closed loop solution is obtained in terms of a set Q_N which is defined in $\mathcal{C}$-space. For arbitrary N, this set has been obtained only for plants with integration and underdamped plants with tuning. However, known properties of the input sequence in the set F_N and on the boundary of the set Γ_N may be of help in obtaining Q_N for other second order plants when $N > 4$.

II. FIRST ORDER SYSTEMS

It was shown in Chapter II that for the first order plant given by

$$G_p(s) = \frac{1}{s + \lambda} \quad , \tag{4-1}$$

the unconstrained minimum fuel sequence is given by Equations (2-74) through (2-77). Consider the case $\lambda > 0$. The optimum fuel input sequence is, from Equation (2-74),

$$u(1) = u(2) = \dots = u(N-1) = 0, \; u(N) = \underline{c}/e^{(N-1)\lambda T} \quad . \tag{4-2}$$

Without any loss in generality, let the initial state $\underline{c}$ be positive. Then if

$$\underline{c} > e^{(N-1)\lambda T} \quad , \tag{4-3}$$

the last member, u(N), exceeds the saturation limit, so that, although the sequence of Equation (4-2) does of course satisfy the deadbeat constraint

$$\underline{c} = \sum_{j=1}^{N} u(j) \; \underline{h}_j \tag{4-4}$$

$$= \sum_{j=1}^{N} u(j) \; e^{(j-1)\lambda T} \quad , \tag{4-5}$$

it does not satisfy the saturation constraint

$$|u(j)| \leqslant 1, \quad j = 1, 2, \dots, N \quad . \tag{4-6}$$

Let the set of all initial states whose linear fuel optimum input sequence satisfies the saturation constraint be called F_N. For first order systems F_N is a portion of the real line, given by

$$F_N = \left(\underline{c} \; \middle| \; \underline{c} = u(N) e^{(N-1)\lambda T}; \quad |u(N)| \leqslant 1 \right) \quad . \tag{4-7}$$

If $\underline{c}$ does not lie in F_N, but does lie in Γ_N, where for first order systems,

$$\Gamma_N = \left(\underline{c} \;\middle|\; \underline{c} = \sum_{j=1}^{N} u(j) e^{(j-1)\lambda T}; \quad |u(j)| \leqslant 1, \; j = 1, 2, \ldots, N \right), \tag{4-8}$$

then there is, by definition, a sequence satisfying Equations (4-5) and (4-6). The problem of finding the input sequence which minimizes

$$F = \sum_{j=1}^{N} |u(j)| \tag{4-9}$$

subject to these equations can be solved intuitively for first order systems. The telescoping rod analogy, described in Chapter III, can be again used to advantage.

There are available N rods, corresponding to the invariant vectors $\underline{h}_j$, $j = 1, 2, \ldots, N$, able to be extended continuously from zero length up to a maximum length $e^{(j-1)\lambda T}$, $j = 1, 2, \ldots, N$. For first order systems these same rods are to be placed end to end along the initial condition line, beginning at the origin and stretching out to reach the initial state. The fuel used is measured by summing the fractional extension of each rod used.

The maximum length of the N-th rod, corresponding to $\underline{h}_N$, is greater than any of the others, so that if $\underline{c}$ is not in F_N, this rod should remain at its maximum length, $|u(N)| = 1$, and the task of reaching $\underline{c}$ continued with the next longest rod, corresponding to $\underline{h}_{N-1}$. If

$$|\underline{c}| > e^{(N-1)\lambda T} + e^{(N-2)\lambda T} \tag{4-10}$$

the (N-2)-th rod is used, and this process continues until $\underline{c}$ is eventually reached. For N = 3, Figure 53 shows the members of the fuel optimum input sequence, and the sets Γ_3 and F_3. The optimum sequence is denoted $\underline{u}^f$. Figure 53 is essentially a graphical method of finding the input sequence $\underline{u}^f$ in an open loop manner.

If $\lambda = 0$, the plant is given by

$$G_p(s) = \frac{1}{s} \quad , \tag{4-11}$$

and each invariant vector is of unit length. The sequence $\underline{u}^f$ is, therefore, not unique, unless of course $\underline{c}$ lies on the tip of Γ_N. The set Γ_N is simply the set of all initial states $\underline{c}$ that satisfy

$$-N \leqslant \underline{c} \leqslant N \quad . \tag{4-12}$$

Choosing the input sequence of Equation (2-75), that is,

$$u(j) = \underline{c}/N, \quad j = 1, 2, \ldots, N, \tag{4-13}$$

the set F_N is equal to the set Γ_N. Therefore, if, from Equation (4-13), $|u(j)| > 1$, then $\underline{c}$ is not in Γ_N and no solution is possible.

The input sequence for the case $\lambda < 0$, corresponding to an unstable plant, is obtained in an open loop manner by exactly the same considerations used for the case $\lambda > 0$. The essential difference is that, since $\underline{h}_1$ is now the longest vector, F_N is always the set of initial states $\underline{c}$ satisfying

$$|\underline{c}| \leqslant 1 \ . \tag{4-14}$$

The set Γ_N does not increase indefinitely with increasing N, but rather approaches a limit:

$$\lim_{N \to \infty} \Gamma_N = \left(\underline{c} \;\middle|\; \underline{c} = \frac{1}{1 + e^{\lambda T}} \ , \lambda < 0 \right) . \tag{4-15}$$

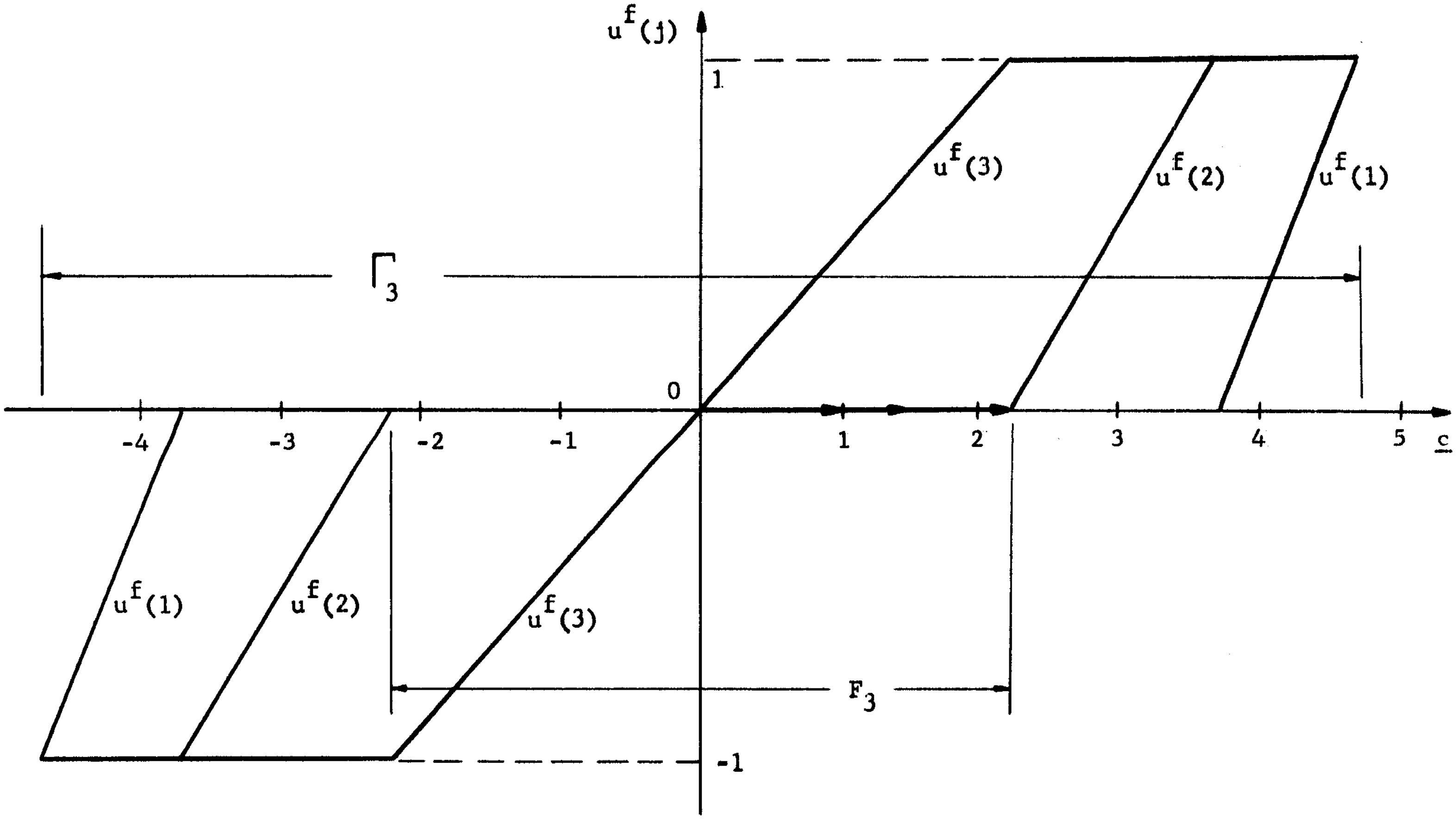


Figure 53. The fuel optimum control sequence for a first order plant with amplitude constrained inputs and N = 3.

III. HIGHER ORDER SYSTEMS

For the general n-th order system, the minimum fuel problem with amplitude constrained inputs is the problem of minimizing Equation (4-9) subject to both Equation (4-6) and the deadbeat constraint. In the state space, the deadbeat constraint is

$$C\,\underline{u} = \underline{x}(0) \quad , \tag{4-16}$$

and in $\mathcal{C}$ -space the constraint is

$$\underline{c} = \underline{a} + H\,\underline{b} \quad . \tag{4-17}$$

In the solution space the set of sequences that satisfy Equations (4-6) and (4-16) is the intersection of the (N-n)-dimensional hyperplane of Equation (4-16) with the N-dimensional hypercube defined by Equation (4-6). Assuming that $\underline{x}(0)$ is in Γ_N, Equation (A-60), the minimum fuel problem is to find a point $\underline{u}$ in this intersection which minimizes the fuel, Equation (4-9). Any point $\underline{u}$ in the solution space is associated with a certain fuel consumption, just as it is associated with a certain energy consumption: see, for example, Figure 28, page 75. Figure 54 shows, for the case N = 2, the iso-fuel surfaces for three different values of F. Figure 54 also shows the set of u(1) and u(2) which satisfy the deadbeat constraint of a first order system. This is the line

$$\underline{x}(0) = u(1)\,\underline{r}_1 + u(2)\,\underline{r}_2 \quad , \tag{4-18}$$

where $\underline{x}(0)$ and the first two canonical vectors, $\underline{r}_1$ and $\underline{r}_2$, are scalars. The minimum fuel input sequence without the saturation constraint is then u(1) = 0, u(2) $\approx$ 1.45, and the constrained minimum fuel sequence is

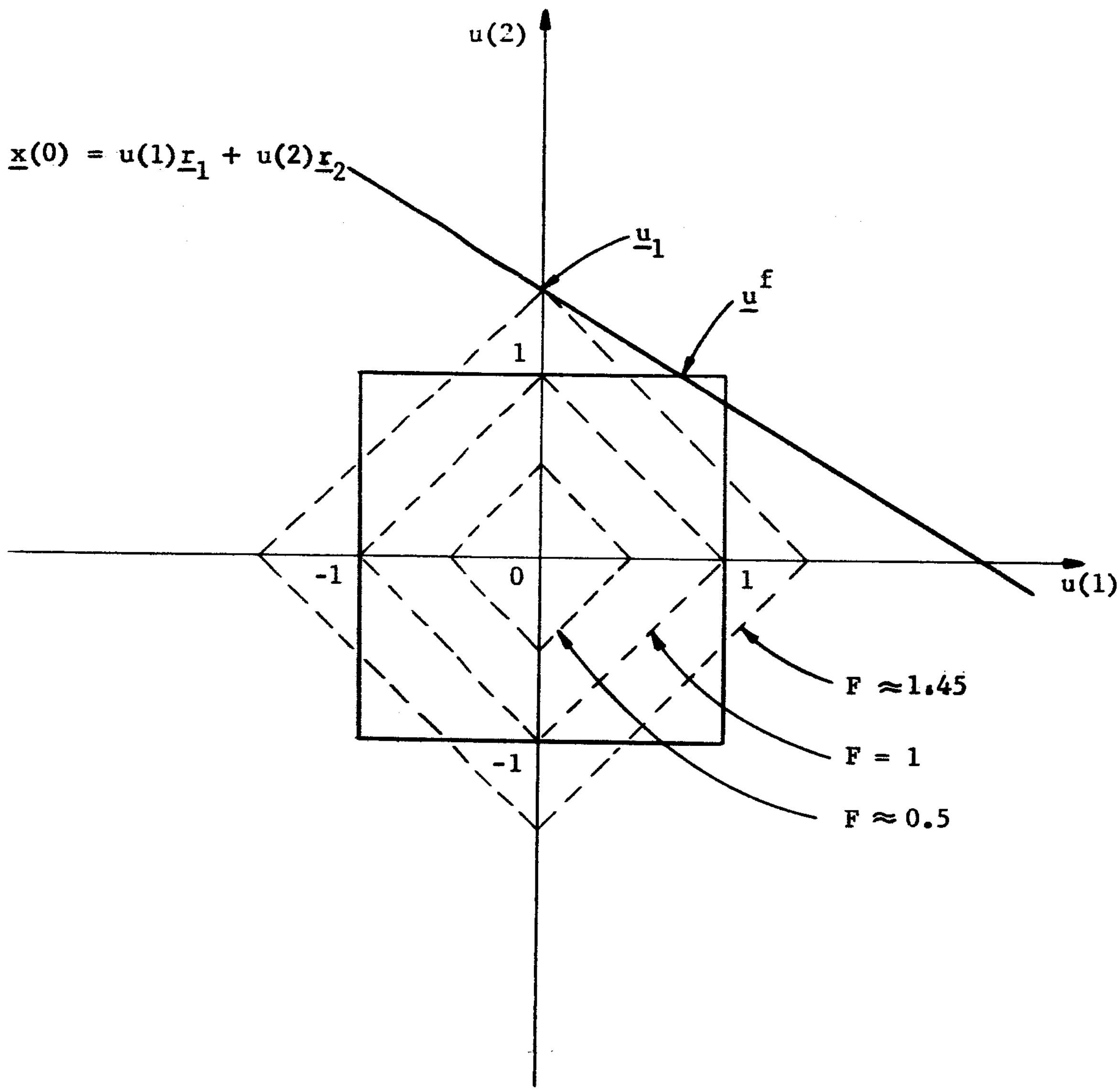


Figure 54. Iso-fuel surfaces for a first order plant with N = 2, giving the linear, $\underline{u}_1$, and constrained, $\underline{u}^f$, fuel optimum input sequences.

$u^f(1) = 1$, $u^f(1) \approx 0.75$. These points are shown as $\underline{u}_1$ and $\underline{u}^f$ respectively in Figure 54.

The general problem can be formulated as a linear programming problem, and Torng gave a detailed example to illustrate the method of formulation and solution (29). However, the method requires the use of a digital computer. In order to gain insight into the minimum fuel problem, techniques are developed similar to those used in the minimum energy problem.

The Fuel Problem in β-space

β-space can be utilized to consider the constrained minimum fuel input sequence. Since U_{N-n} is the set of all corrections $\underline{\beta}$, which when added to $\underline{b}^o$ give an input sequence satisfying Equations (4-6) and (4-17), this set not only contains the correction for the minimum energy input sequence, but also contains at least one correction, $\underline{\beta}^f$, which gives a solution to the minimum fuel problem. Let such an input sequence be called $\underline{u}^f$. Then

$$\underline{u}^f = \underline{u}^o + \underline{\delta}^f \tag{4-19}$$

where $\underline{u}^o$ is the linear minimum energy input sequence, and $\underline{\delta}^f$ is given by Equations (3-47) and (3-54) as

$$\underline{\delta}^f = \begin{bmatrix} \underline{\alpha}^f \\ \\ \underline{\beta}^f \end{bmatrix} = \begin{bmatrix} -H\,\underline{\beta}^f \\ \\ \underline{\beta}^f \end{bmatrix} . \tag{4-20}$$

Before considering further how to obtain $\underline{\beta}^f$, it should be noted that when β-space is used, it becomes necessary to calculate $\underline{u}^o$. While

$\mathcal{B}$-space may be useful to consider the fuel problem, for example, if it is desired to compare $\underline{\beta}^e$ and $\underline{\beta}^f$, the calculation of $\underline{u}^o$ may be avoided by studying the fuel problem in the partitioned solution space.

The Fuel Problem in the Partitioned Solution Space

By partitioning the matrix C into R and Q and the input sequence $\underline{u}$ into $\underline{a}$ and $\underline{b}$, the deadbeat constraint of Equation (4-16) was transformed into the deadbeat constraint of Equation (4-17). Just as the correction space containing the N-vector $\underline{\delta}$ was partitioned into α-space and β-space, each containing respectively the n-vector $\underline{\alpha}$ and the (N-n)-vector $\underline{\beta}$, the solution space may be partitioned into two spaces: $\mathcal{A}$-space being n-dimensional with coordinates u(1), ..., u(n) containing the vector $\underline{a}$, and $\mathcal{B}$-space being (N-n)-dimensional with coordinates u(n+1), ..., u(N) containing the vector $\underline{b}$. The relationship between these spaces $\mathcal{A}$ and $\mathcal{B}$ is then given by Equation (4-17).

If the input sequence is to satisfy the saturation constraint, the components of $\underline{a}$ and $\underline{b}$ must satisfy Equation (4-6). Let the set A in $\mathcal{A}$-space be the set of all $\underline{a}$ such that

$$|u(j)| \leqslant 1, \quad j = 1, 2, \ldots, n, \tag{4-21}$$

and let the set B in $\mathcal{B}$-space be the set of all $\underline{b}$ that satisfy

$$|u(j)| \leqslant 1, \quad j = n+1, \ldots, N. \tag{4-22}$$

These sets are respectively n-dimensional and (N-n)-dimensional hypercubes, centered on the origins of their respective spaces. Now consider the deadbeat constraint, Equation (4-17). Assume that $N \geqslant 2n$. To any initial state $\underline{c}$ and input $\underline{a}$ correspond points in $\mathcal{B}$-space lying on the (N-n)-

dimensional hyperplane,

$$H \underline{b} = \underline{c} - \underline{a} \,. \tag{4-23}$$

If $N = 2n$, the hyperplane reduces to a single point, since H is then $n \times n$ and may be inverted. The vector $\underline{a}$ must lie in the set A. Therefore, the set of $\underline{b}$ which satisfy Equations (4-17) and (4-21) is the map of the set of points $\underline{c} - \underline{a}$, for all $\underline{a}$ in the set A, from $\mathcal{A}$-space into $\mathcal{B}$-space. Let this set of $\underline{b}$ be called A'. Let the intersection of B and A' be called U. Then if $\underline{c}$ is in Γ_N, U contains at least one point $\underline{b}$ such that the input sequence

$$\underline{u} = \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix} \tag{4-24}$$

satisfies the deadbeat and saturation constraints. The similarity between $\mathcal{B}$-space and β-space is evident. In fact, if to any point in $\mathcal{B}$-space is added the vector $-\underline{b}^o$, and the coordinates of $\mathcal{B}$-space are changed to those of β-space, the sets B_{N-n}, A'_n and U_{N-n} are respectively identical to B, A' and U.

Let the fuel cost associated with $\underline{a}$ be F_A, then

$$F_A = |u(1)| + \ldots + |u(n)| \,, \tag{4-25}$$

and that associated with $\underline{b}$ be F_B, then

$$F_B = |u(n+1)| + \ldots + |u(N)| \,. \tag{4-26}$$

The total fuel consumption is therefore $F = F_A + F_B$. The minimum fuel problem with input saturation amounts to finding the point $\underline{b}$ in U which minimizes $F_A + F_B$. The linear solution, which of course need not be in

U, lies on at least one of the hyperplanes $u(j) = 0$, $j = 1, 2, \ldots, N$, since $N > n$ and a linear fuel optimum input sequence can be given with at least N-n members of the sequence equal to zero. An optimum solution under the saturation constraint must lie in U. If a linear solution cannot be found in U, the constrained solution must lie on the boundary of U. To illustrate this formulation in $\mathcal{B}$-space, Figure 55 shows the situation for a typical second order system with $N = 4$. The iso-fuel lines F_B are partially shown by the dashed lines. Each iso-fuel line shown, for both F_A and F_B, is separated from the next by an increment $F = 0.1$. The linear solution is found by starting at any point in the space and moving so that the sum $F = F_A + F_B$ is reduced. A minimum fuel solution is obtained at a particular point $\underline{b}$ when any other point in its neighbourhood causes $F_A + F_B$ to increase. In the example, the solution lies at the crossing of the lines $u(2) = 0$ and $u(3) = 0$. This point, which is unique and is marked with a small circle for clarity, is not in the set U. The optimum unique constrained solution lies at the intersection of the lines $u(2) = 0$ and $u(1) = 1$, and is also shown encircled.

While partitioning the solution space in itself provides a graphical solution to the minimum fuel problem only for the cases $n = 1$, $N \leqslant 3$, and $n = 2$, $N \leqslant 4$, it is useful as a means to investigate the properties of the minimum fuel solution, just as β-space was used to visualize the properties of the minimum energy problem.

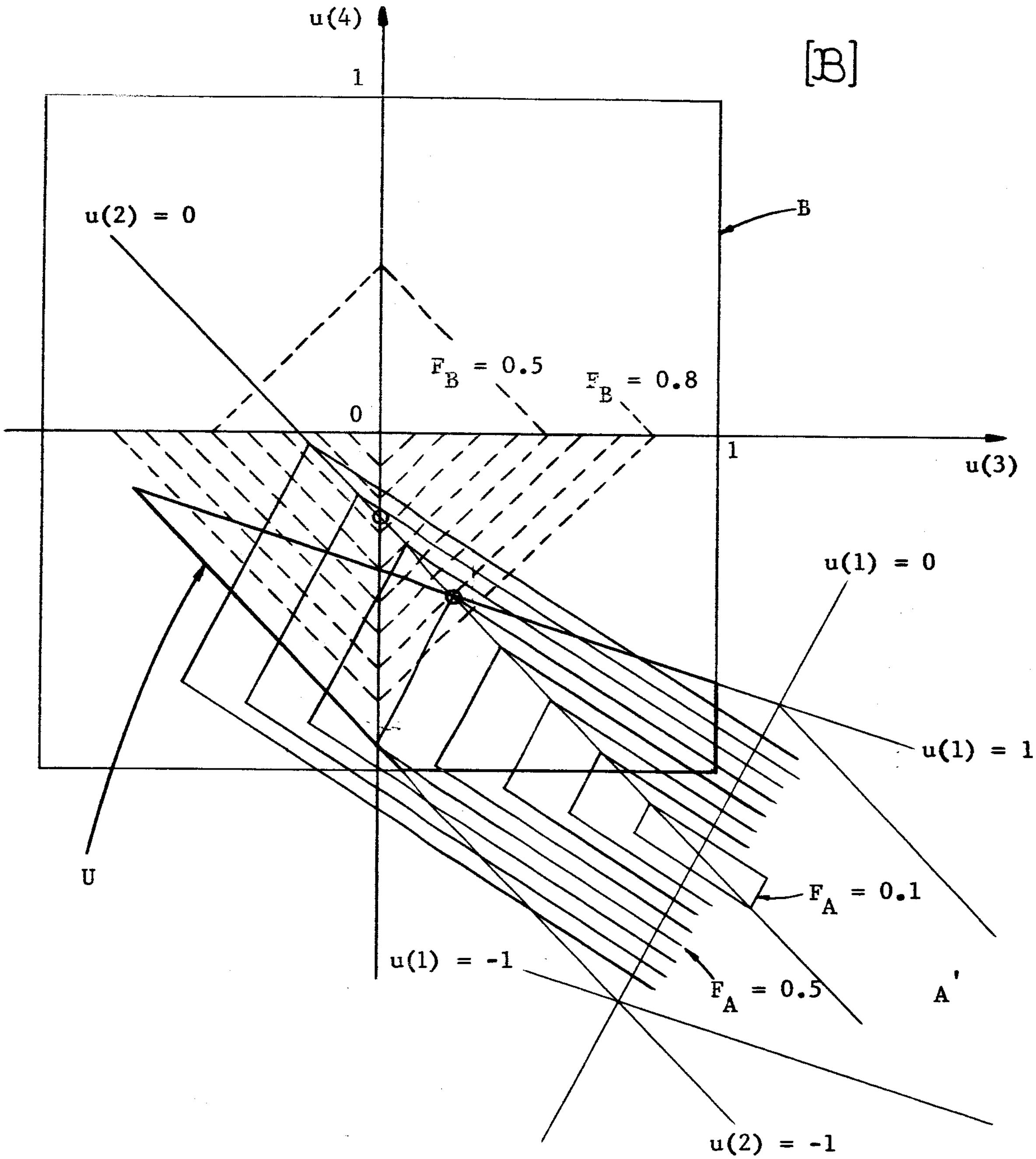


Figure 55. The minimum fuel problem solved in $\mathcal{B}$-space for a second order system with N = 4.

IV. SECOND ORDER SYSTEMS

As in the minimum energy problem, it is of interest to know for what initial states the linear minimum fuel input sequence satisfies the saturation constraints, and is therefore itself a solution to the constrained problem.

The Set F_N

Let the set of all initial states whose linear minimum fuel input sequence satisfies the saturation constraints be called F_N. The linear fuel optimum sequence, as shown in Chapter II, is not necessarily unique. In order that the set F_N have meaning, one method of generating the input sequence must be established. For the moment, however, consider the case where this uniqueness problem does not arise.

Recall that the faces of $S_N(1)$ are the 2p line segments $L_s(\pm i,\pm j)$, and these segments generate the cones $C_s(\pm i,\pm j)$ as shown in Figure 11, page 41. For example, the line segment $L_s(i,j)$ is the set of points $\underline{c}$ where

$$\underline{c} = \mu_i \underline{h}_i + \mu_j \underline{h}_j, \quad \mu_i, \mu_j \geqslant 0, \quad \mu_i + \mu_j = 1, \tag{4-27}$$

and the line segment $L_s(-i,j)$ is given by the set of points $\underline{c}$ where

$$\underline{c} = -\mu_i \underline{h}_i + \mu_j \underline{h}_j, \quad \mu_i, \mu_j \geqslant 0, \ \mu_i + \mu_j = 1. \tag{4-28}$$

Suppose $\underline{c}$ lies in $C_s(\pm i,\pm j)$. The optimum input sequence is obtained from Equations (2-84) and (2-85):

$$\underline{c} = u^f(i)\ \underline{h}_i + u^f(j)\ \underline{h}_j, \quad u^f(k) = 0, \quad k \neq i,j. \tag{4-29}$$

Assume for the moment that this gives a unique sequence. This requires that the line segment which ends on $\pm \underline{h}_i$, the line segment $L_s(\pm i,\pm j)$ and the line segment which starts from $\pm \underline{h}_j$ do not lie on the same straight line. Then the set of all $\underline{c}$ given by Equation (4-29) which give an input sequence satisfying the saturation constraint is given by

$$\underline{c} = \pm u(i)\ \underline{h}_i \pm u(j)\ \underline{h}_j, \quad 0 \leq u(i) \leq 1, \quad 0 \leq u(j) \leq 1. \tag{4-30}$$

For example, if $\underline{c}$ lies in $C_s(i,j)$ the set is given by all $\underline{c}$ satisfying

$$\underline{c} = +u(i)\ \underline{h}_i + u(j)\ \underline{h}_j, \quad 0 \leq u(i) \leq 1, \quad 0 \leq u(j) \leq 1, \tag{4-31}$$

and if $\underline{c}$ lies in the cone $C_s(-i,j)$ the set is given by all $\underline{c}$ satisfying

$$\underline{c} = -u(i)\ \underline{h}_i + u(j)\ \underline{h}_j, \quad 0 \leq u(i) \leq 1, \quad 0 \leq u(j) \leq 1. \tag{4-32}$$

If the linear input sequence is unique for all initial states in $\mathcal{C}$, the set F_N is given uniquely. For each of the 2p cones $C_s(\pm i,\pm j)$, form the set of all $\underline{c}$ satisfying Equation (4-30). This set is the set F_N. F_N is of course symmetric with respect to the origin and contains the set $S_N(1)$. An example of this set is given in Figure 56, for a typical underdamped second order plant with N = 5.

Now consider the case where there is a region in Γ_N for which $\underline{c}$ has no unique optimum input sequence. For example, the plant $1/s^2$, see Figure 14, page 47, has the regions C(1,N) and C(-1,-N) in which this problem of non-uniqueness arises. One way to get around this problem would be to use Equation (4-29) regardless of whether the sequence is unique or not, and then F_N is given uniquely. This is not very satisfactory however, since another set F_N can always be found

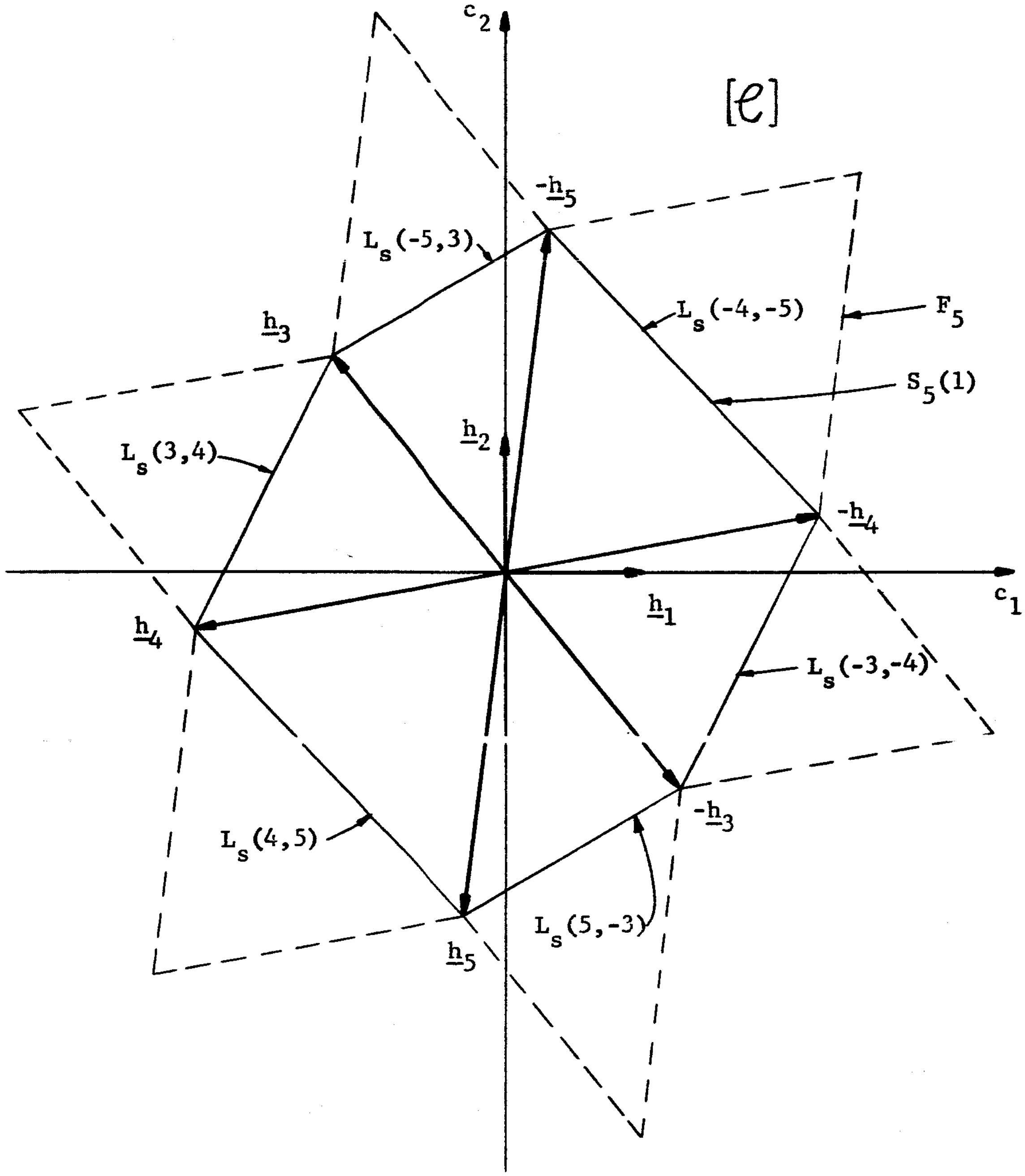


Figure 56. The set $S_5(1)$ and the corresponding set F_5 for a typical second order underdamped plant.

merely be choosing another rule for states having non-unique optimum input sequences. However, once one particular method of choosing the input sequence has been decided upon, the set F_N can be generated.

The Fuel Problem in $\mathcal{A}$-space

If $N \leqslant 2n$, the fuel problem can be conveniently formulated in $\mathcal{A}$-space. The deadbeat constraint, Equation (4-17), gives

$$\underline{a} = \underline{c} - H\,\underline{b} \quad . \tag{4-33}$$

Equation (4-33) may be written as

$$\underline{a} = \underline{c} - \underline{h}_{n+1}\, u(n+1) - \underline{h}_{n+2}\, u(n+2) - \ldots - \underline{h}_N\, u(N) \quad . \tag{4-34}$$

The set,

$$\Gamma'_{N-n} = \left(\underline{a} \,\middle|\, \underline{a} = \sum_{j=n+1}^{N} \underline{h}_j\, u(j); \quad |u(j)| \leqslant 1, \ j = n+1, \ldots, N \right) , \tag{4-35}$$

is an n-dimensional polygon centered at the origin of $\mathcal{A}$-space. The right hand side of Equation (4-34), under the saturation constraints of Equation (4-22) is therefore the set $-\Gamma'_{N-n}$ centered on the point $\underline{a} = \underline{c}$. With $N \leqslant 2n$, any point $\underline{a}$ in A uniquely defines an entire input sequence, and, if the point also lies in the polygon $\underline{c} - \Gamma'_{N-n}$, the sequence will take $\underline{c}$ into the origin while satisfying the saturation constraints. The fuel consumption of such a sequence is $F = F_A + F_B$.

Second order systems, $n = 2$, can be examined graphically for $N = 3$ and $N = 4$. Although these cases are very restricted, they do reveal several interesting aspects of the minimum fuel problem, and indicate how more general cases might be approached.

<u>The case N = 3</u>. The set A is given by Equation (4-21) as a square centered on the origin, and the right hand side of Equation (4-34), satisfying Equation (4-22), is the set of all $\underline{a}$ given by,

$$\underline{a} = \underline{c} - \underline{h}_3\, u(3), \quad |u(3)| \leqslant 1. \tag{4-36}$$

Therefore, if $\underline{a} = \underline{c}$, $F_B = 0$, and if $\underline{a} = \underline{c} \pm \underline{h}_3$, $F_B = 1$. The point $\underline{a}$ which minimizes the fuel, $F = F_A + F_B$, can therefore be found for any state $\underline{c}$ in Γ_3. The coordinates of $\mathcal{C}$-space serve to give $\underline{a}$ directly, since when $u(3) = 0$, then $\underline{a} = \underline{c}$. Thus, any $\underline{c}$ in $\mathcal{C}$ gives the vector $\underline{a}$ directly, which in turn gives the vector $\underline{b}$. Figure 57 shows a typical second order system, which, as shown for N = 3, gives a unique fuel optimum input sequence for any initial state. The set F_N is shown by the heavily dashed line, and the iso-fuel lines F_A as the lightly dashed lines. The following properties of the minimum fuel input sequence with input saturation can be observed for this example.

1. In region abcd, $u^f(1) = 1$.
2. In region ecgh, $u^f(3) = -1$.
3. In region oegh, $u^f(1) = 0$.
4. In region phgm, $u^f(2) = -1$.
5. In region koph, $u^f(3) = 0$.

Because of symmetry these regions are sufficient to characterize the sequence for all states in Γ_3, since if one of the inputs is fixed, the other two are given by Equation (4-36). This method of obtaining the input sequence is not very convenient, especially since the initial state must be identified as belonging to a particular region, and

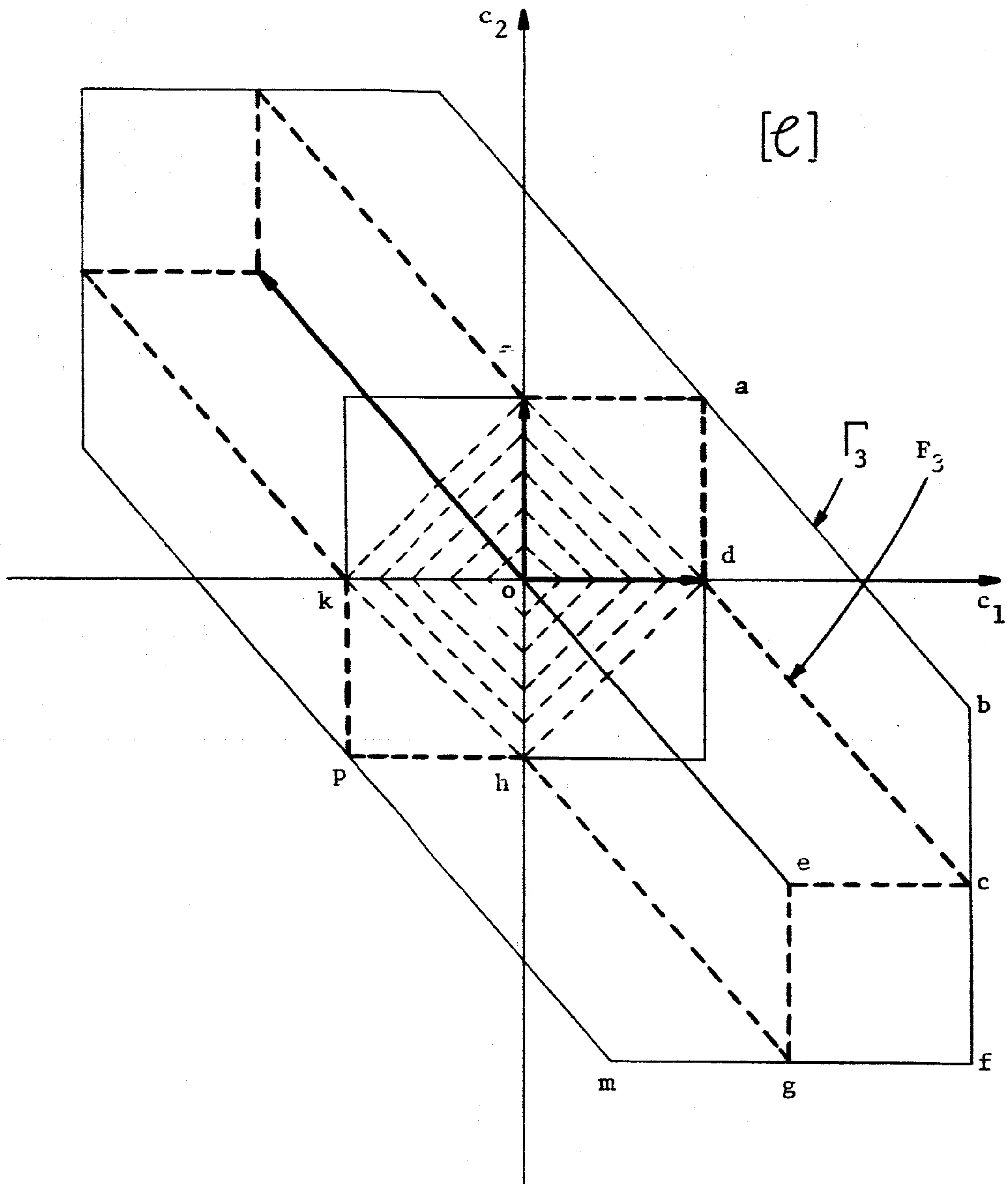


Figure 57. The sets Γ_3 and F_3 divided into regions characterizing the fuel optimum input sequence.

furthermore, the situation becomes more complicated for N = 4. A closed loop method results in a much simpler way of obtaining the input sequence. A closed loop method requires that, given a settling time of N sampling periods, only the first input be determined. The problem is then repeated for a settling time of N-1 sampling periods and so on, until N = 1. The controller then identifies the state at each stage of the regulation process, generating in turn u(1) for an N-member input sequence, then u(1) for an (N-1)-member input sequence, and so on until N = 1, when obtaining u(1) for N = 1 completes the regulation.

In Figure 57, the region oegh contains initial states whose three-member fuel optimum input sequence has $u^f(1) = 0$. In abcd, $u^f(1) = 1$. Figure 58 shows these two regions and their symmetrical counterpart. The regions $|u^f(1)| = 1$ have horizontal cross-hatching, and those for $u^f(1) = 0$ have vertical cross-hatching. In general, let the set of all states where $u^f(1) = 0$ for a given N, be called Q_N. Therefore, the region in Figure 58 with the vertical cross-hatching is denoted Q_3. The regions in Γ_3 between the cross-hatched regions contain states where $|u^f(1)| \leq 1$. Suppose the initial state is given in a region where $|u^f(1)| < 1$, and consider projecting a line parallel to $\underline{h}_1$ from the point $\underline{c}$ until it just touches the region Q_3. The length of this line defines $|u^f(1)|$. If, in order to reach Q_3, the projection is in the direction $-\underline{h}_1$, then $u^f(1) > 0$. If the projections needs to be in the direction $+\underline{h}_1$, then $u^f(1) < 0$. This can be formalized as follows. Assume $\underline{c}$ is in Γ_3 and let $\underline{c} - \mu\underline{h}_1$ lie on the boundary of Q_3. Then

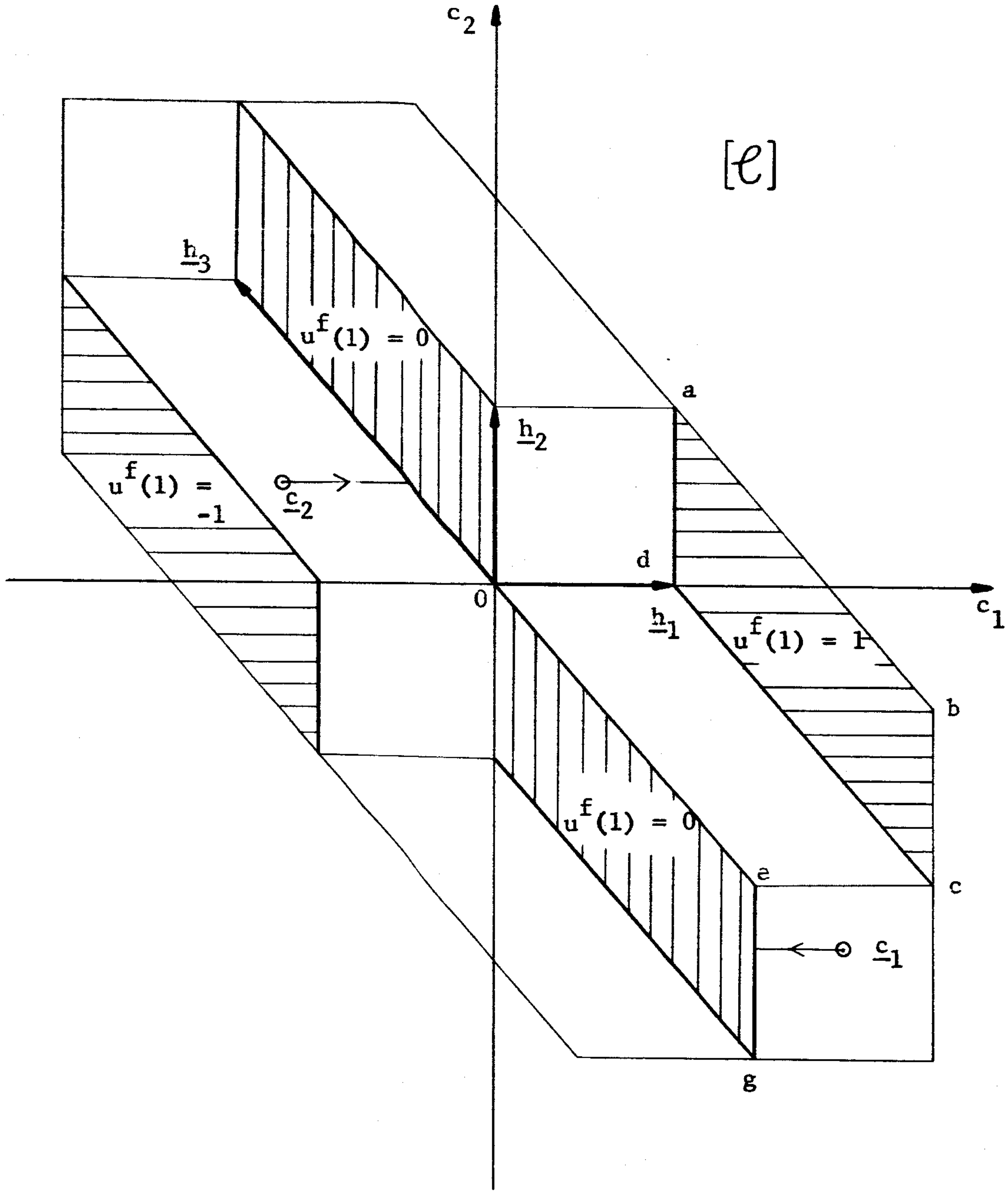


Figure 58. Characterization of $u^f(1)$ for N = 3.

$$\text{if } \underline{c} \text{ is not in } Q_3, \quad u^f(1) = \text{sat. } \mu \ ,$$

$$\text{if } \underline{c} \text{ is in } Q_3, \quad u^f(1) = 0 \ , \tag{4-37}$$

where the "sat." function is defined as

$$\text{sat. } \mu = \begin{cases} -1 & \text{if } \mu \quad -1 \\ \mu & \text{if } |\mu| < 1 \\ 1 & \text{if } \mu \quad 1 \end{cases} \ . \tag{4-38}$$

Figure 58 shows two initial states $\underline{c}_1$ and $\underline{c}_2$. For $\underline{c}_1$, $u^f(1) \approx 0.5$, and for $\underline{c}_2$, $u^f(1) \approx -0.75$. Having generated and applied the first input for N = 3, the initial state will have moved, assuming no adverse disturbances, from $\underline{c}$ to $\underline{c}'$ in Γ_2. In Γ_2 the remainder of the sequence is uniquely determined by

$$\begin{bmatrix} u^f(2) \\ u^f(3) \end{bmatrix} = \begin{bmatrix} c_1' \\ c_2' \end{bmatrix} \ . \tag{4-39}$$

Figures 59, 60, 61 and 62 show representative examples of the set Q_3 and its relation to Γ_3. Figure 59 illustrates Q_3 and Γ_3 for the plant

$$G_p(s) = \frac{1}{s^2} \ . \tag{4-40}$$

Figure 60 corresponds to the plant

$$G_p(s) = \frac{1}{s^2 + 2as + b^2} \ , \tag{4-41}$$

where the poles, real or complex, satisfy the non-uniqueness criterion of Equation (2-91). Figure 61 gives Q_3 for the plant

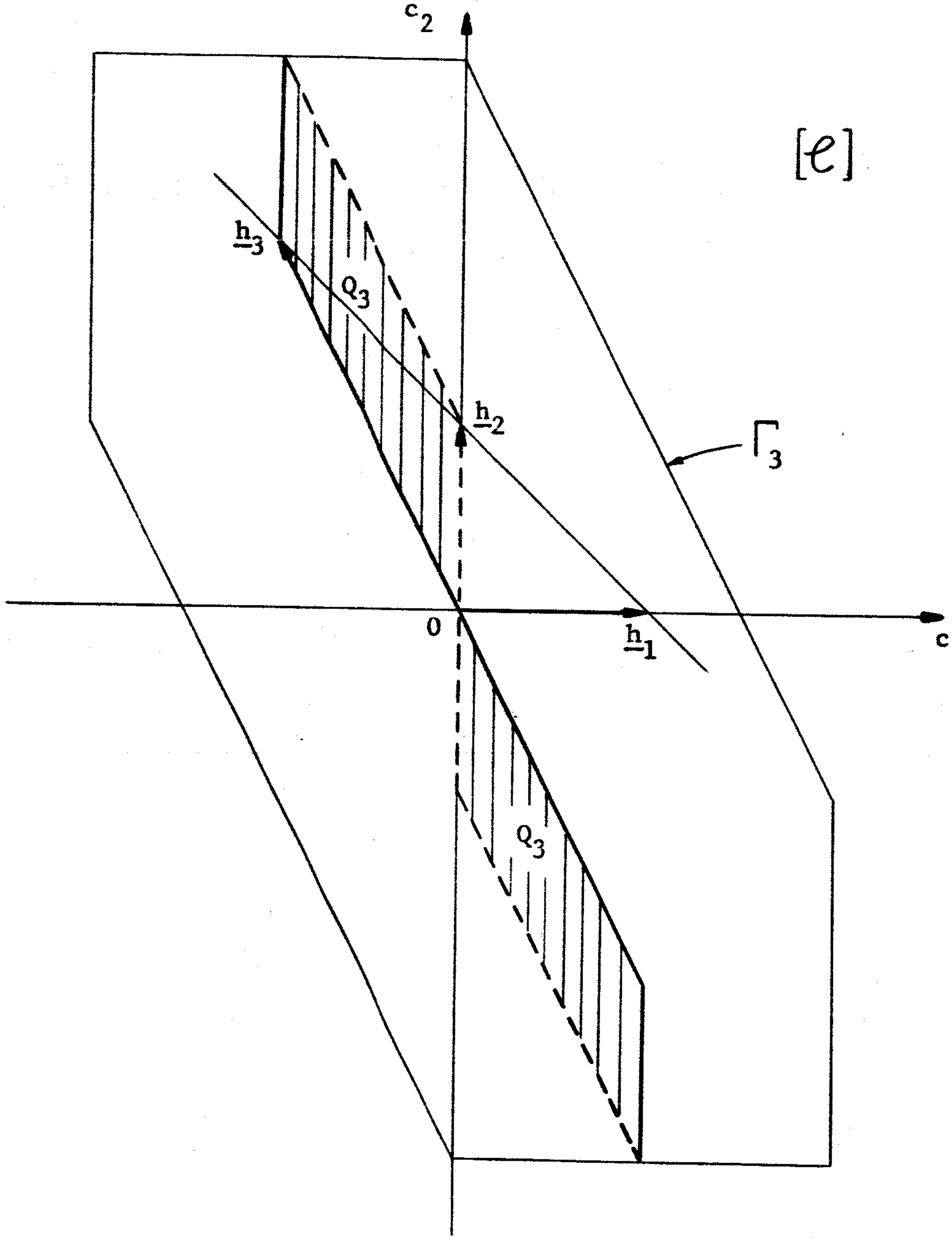


Figure 59. The set Q_3 for the plant $1/s^2$.

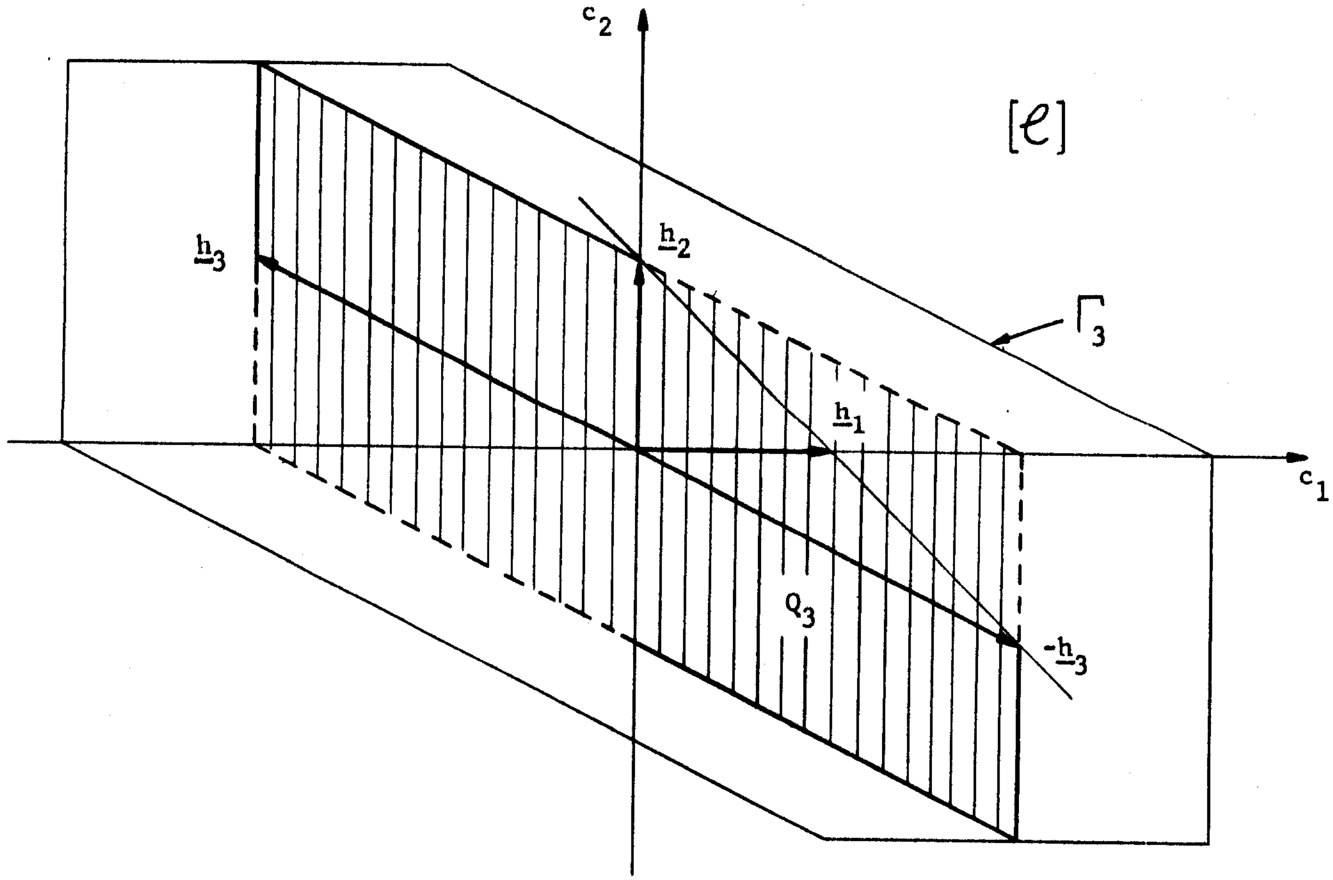


Figure 60. The set Q_3 for the plant of Equation (4-41).

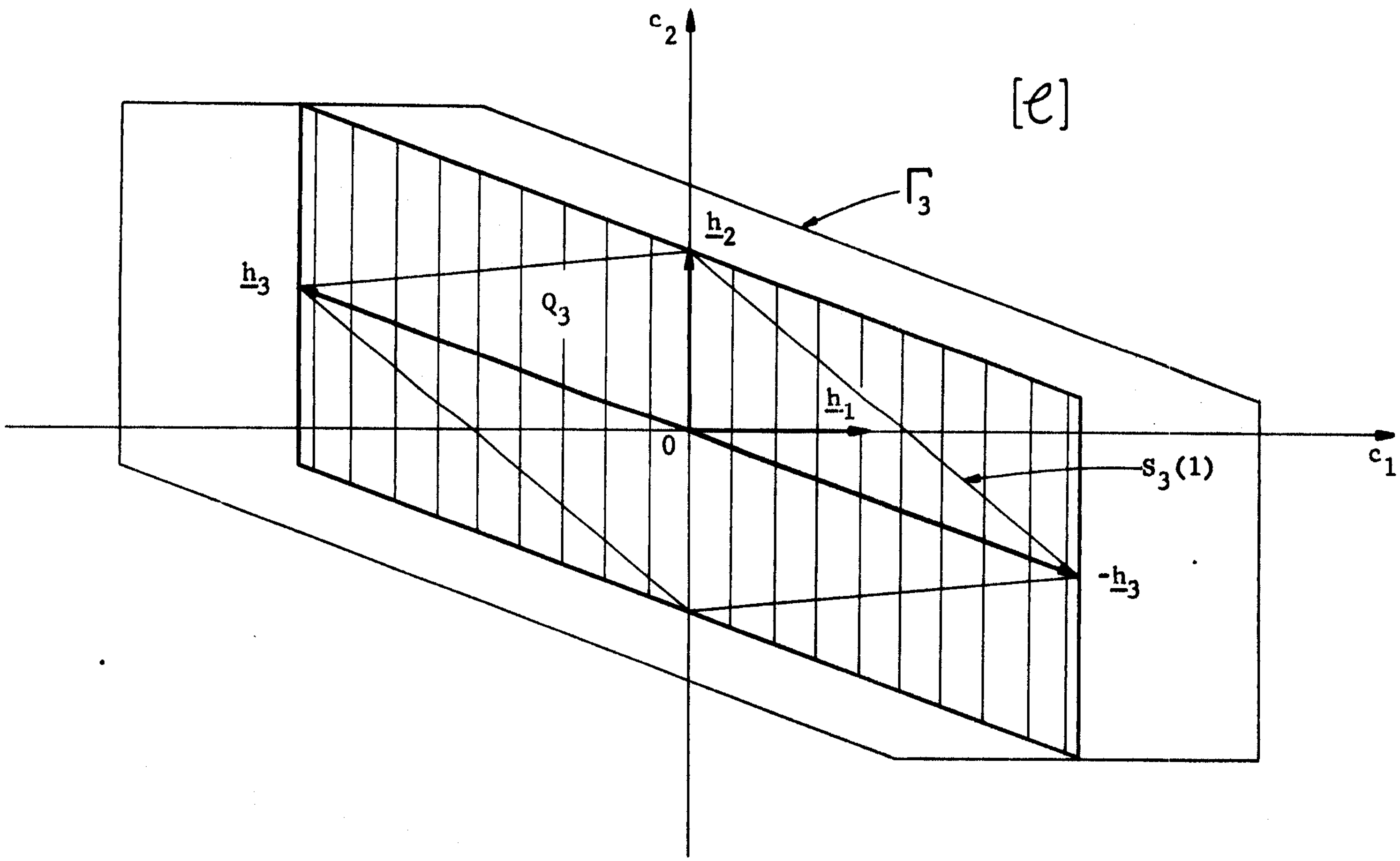


Figure 61. The set Q_3 for the plant of Equation (4-42).

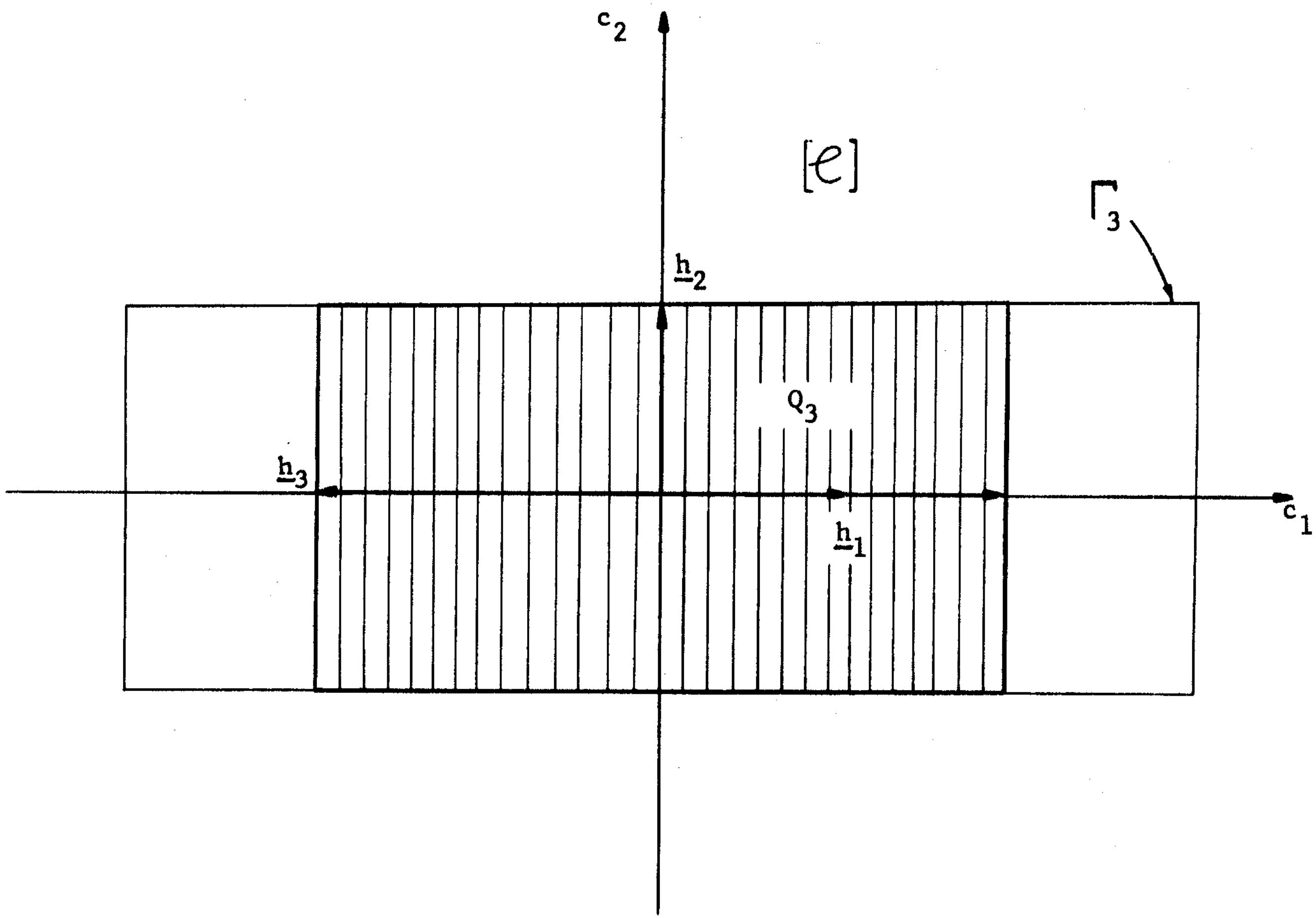


Figure 62. The set Q_3 for a tuned underdamped second order plant.

$$G_p(s) = \frac{1}{(s + a + jb)(s + a - jb)} \quad , \tag{4-42}$$

where the poles are such that the input sequence is always unique, and finally, Figure 62 shows a plant of the form of Equation (4-42) where the poles have been tuned.

Consider the plant of Equation (4-40). Since it is known, Chapter II, page 46, that the linear minimum fuel sequence is not unique in C(1,3) and C(-1,-3), it is to be expected that Q_3 is not unique. The vertically cross-hatched region of Figure 59, page 155, shows one possible extreme that Q_3 may take. This is the set

$$Q_3 = \left(\underline{c} \;\middle|\; \underline{c} = \sum_{j=2}^{3} \mu_j \, \underline{h}_j, \quad 0 \leq \mu_j \leq 1 \right) \quad . \tag{4-43}$$

The smallest Q_3 is the set of points lying on the line formed by joining $\underline{h}_2$ to $\underline{h}_3$ and $-\underline{h}_2$ to $-\underline{h}_3$. This is part of the boundary of the largest possible Q_3, Equation (4-43), and is shown by the solid line in Figure 59. Another plant which has a non-unique Q_3 is shown in Figure 60, page 156, corresponding to the plant of Equation (4-41). The largest Q_3 is again shown by the vertically cross-hatched region, and is

$$Q_3 = \left(\underline{c} \;\middle|\; \underline{c} = \sum_{j=2}^{3} \mu_j \, \underline{h}_j ; \quad |\mu_j| \leq 1 \right) \quad , \tag{4-44}$$

and the smallest Q_3 is given by

$$Q_3 = \left(\underline{c} \;\middle|\; \underline{c} = \sum_{j=2}^{3} \mu_j \underline{h}_j ; \quad 0 \leqslant \mu_j \leqslant 1 \right) \quad . \tag{4-45}$$

The next two plants have unique sets Q_3. Figure 61, page 157, corresponding to the underdamped plant of Equation (4-42), shows $\underline{h}_1$ interior to the set $S_3(1)$. Figure 62, corresponding to the tuned plant, also has $\underline{h}_1$ interior to the set $S_3(1)$. In both cases Q_3 is given by Equation (4-44). The size of Q_3 relative to Γ_3 is a measure of the usefulness of the first member of the control to the regulation process. If Q_3 is large, as for example when $\underline{h}_1$ is interior to the set $S_3(1)$, see Figure 61 and 62, the main burden of the regulation usually rests upon the last two inputs. When the size of Q_3 is small relative to Γ_3, the first control plays an important role in the regulator process. Figure 59 and, to a lesser extent, Figure 60 are examples where the first input may be at its limit for a large region of initial states in Γ_3.

<u>The case N = 4</u>. It is a little more complicated to obtain the set Q_4, and, depending on the particular invariant vectors of the plant, Q_4 has many more possible shapes than Q_3.

The set A is unchanged, given by Equation (4-21), but the right hand side of Equation (4-34) satisfying Equation (4-22) becomes the set of $\underline{a}$ satisfying

$$\underline{a} = \underline{c} - \left[\underline{h}_3 \, u(3) + \underline{h}_4 \, u(4) \right] \; , \quad |u(3)| \leqslant 1, \quad |u(4)| \leqslant 1. \tag{4-46}$$

The set Q_4 may therefore be found by methods similar to those illustrated in Figure 55, page 145. In practice a transparent (onionskin) template of the polygon

$$\underline{h}_3\, u(3) + \underline{h}_4\, u(4); \quad |u(3)| \leq 1, \quad |u(4)| \leq 1 \tag{4-47}$$

is constructed, and the iso-fuel lines, F_B = constant, are drawn on this template. The iso-fuel lines F_A = constant having also been constructed in the set A, the template may be positioned with its center anywhere in Γ_4, thus examining the first input member of the fuel optimum input sequence for all possible initial states in Γ_4. The plants

$$G_p(s) = \frac{1}{s^2} \tag{4-48}$$

and

$$G_p(s) = \frac{1}{s(s + a)} \tag{4-49}$$

were studied in this manner and the sets Q_4 constructed. Again, as would be expected, for plants with integration the set Q_4 is not unique. Figure 63 shows that the set Q_4 for the plant of Equation (4-48) may vary from just the solid line, formed from $-\underline{h}_4$, $-\underline{h}_3$ $-\underline{h}_2$ and $\underline{h}_4$, $\underline{h}_3$, $\underline{h}_2$, to the entire region enclosed by this solid line and the dashed line. The plant of Equation (4-49) has a set Q_4 formed in exactly the same manner.

For $N > 4$, the problem of finding Q_4 becomes very complicated when tackled in this manner. However, there are two considerations which can help in the graphical construction of the set Q_N.

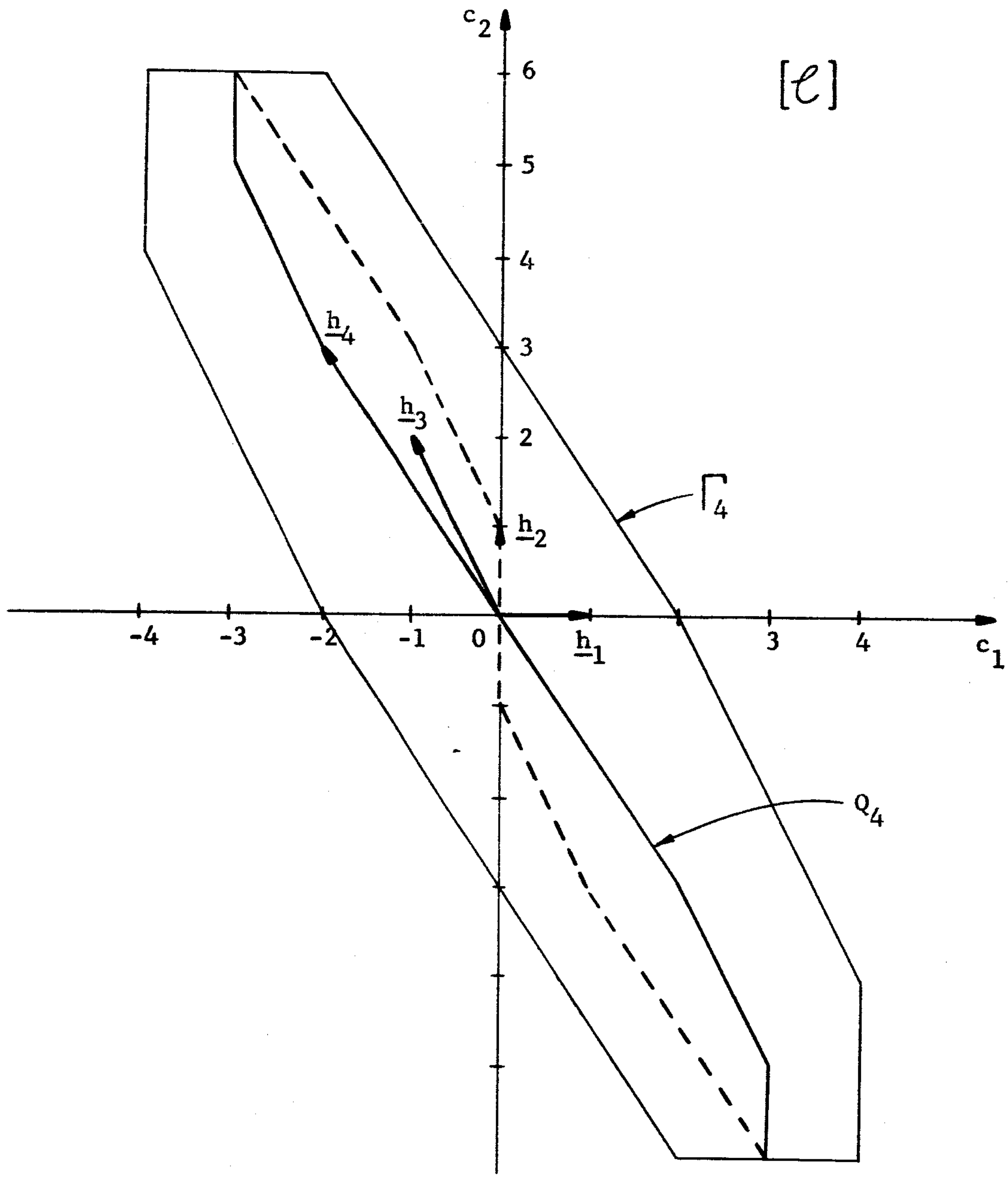


Figure 63. Possible sets Q_4 for the plant $1/s^2$.

1. Initial states lying on the boundary of Γ_N have unique input sequences (the set U, equivalently U_{N-n}, is a point, as is shown for example in Figure 43, page 111). It is a straightforward matter to find the input sequence for initial states on the boundary of Γ_N, and $u(1) = u^f(1)$ can therefore be found uniquely for such states*.

2. The set F_N can be of help in finding Q_N. The set F_N includes a considerable proportion of the states in Γ_N. It can therefore be used to obtain $u^f(1)$ when $\underline{c}$ is in F_N, and to indicate if $u^f(1) = 0$ when $\underline{c}$ is not in F_N.

Second Order Systems with Integration

If the plant has integration; i.e, is of the form of Equation (4-48) or Equation (4-49), the set Q_N can be found for all N. Desoer and Lee (22, page 371) defined the set $T_N(f)$ as the set of all initial states which can be brought to the origin in N sampling periods with a fuel consumption $F \leqslant f$. Therefore, in $\mathcal{C}$-space,

$$T_N(f) = \left(\underline{c} \;\middle|\; \underline{c} = \sum_{j=1}^{N} u(j)\, \underline{h}_j ; \; |u(j)| \leqslant 1, \; \sum_{j=1}^{N} |u(j)| \leqslant f \right) . \tag{4-50}$$

If f = 1, $T_N(f)$ becomes $S_N(1)$, and if f = N, $T_N(f)$ becomes Γ_N. Desoer and Lee demonstrated that $T_N(f)$ is convex and contains the origin, and

---

*There is one exception. If the second order system has tuned complex poles, states on the boundary of Γ_N do not necessarily have unique input sequences. However, as will be shown, for such systems the set Q_N can be found quite easily from other considerations.

that if an initial state $\underline{c}$ is in $\partial T_N(f)$, the boundary of $T_N(f)$, any control sequence which satisfies Equations (4-4) and (4-6), and for which F = f, is an optimum input sequence. After some rather detailed manipulations, it was shown that one set Q_N for second order plants of the form of Equations (4-48) and (4-49) could be given as

$$Q_N = \left(\underline{c} \;\middle|\; \underline{c} = \sum_{j=2}^{N} \mu_j \underline{h}_j ; \quad 0 \leqslant \mu_j \leqslant 1, \; j = 2, 3, \ldots, N \right) . \tag{4-51}$$

Figures 59, page 155, and 63, page 162, have already shown respectively the sets Q_3 and Q_4 given by Equation (4-51). The set Q_N is not unique, and Figure 64 shows alternative sets Q_3 and Q_4 for the plant of Equation (4-48).

Second Order Systems with Tuned Complex Poles

The plant

$$G_p(s) = \frac{1}{(s + a + jb)(s + a - jb)} \tag{4-52}$$

can be tuned by making

$$bT = \frac{\pi}{2} \tag{4-53}$$

as explained in Chapter III, page 130, and the invariant vectors $\underline{h}_1$, $\underline{h}_3$, $\underline{h}_5$, ... lie in $\mathcal{C}$-space on the line $c_2 = 0$, and $\underline{h}_2$, $\underline{h}_4$, $\underline{h}_6$, ... lie on the line $c_1 = 0$. Figure 52, page 131, shows these invariant vectors for a typical tuned plant. The fuel optimum input sequence can be found for these tuned plants in a very straightforward manner, by either open or closed loop methods. The invariant vectors for a tuned

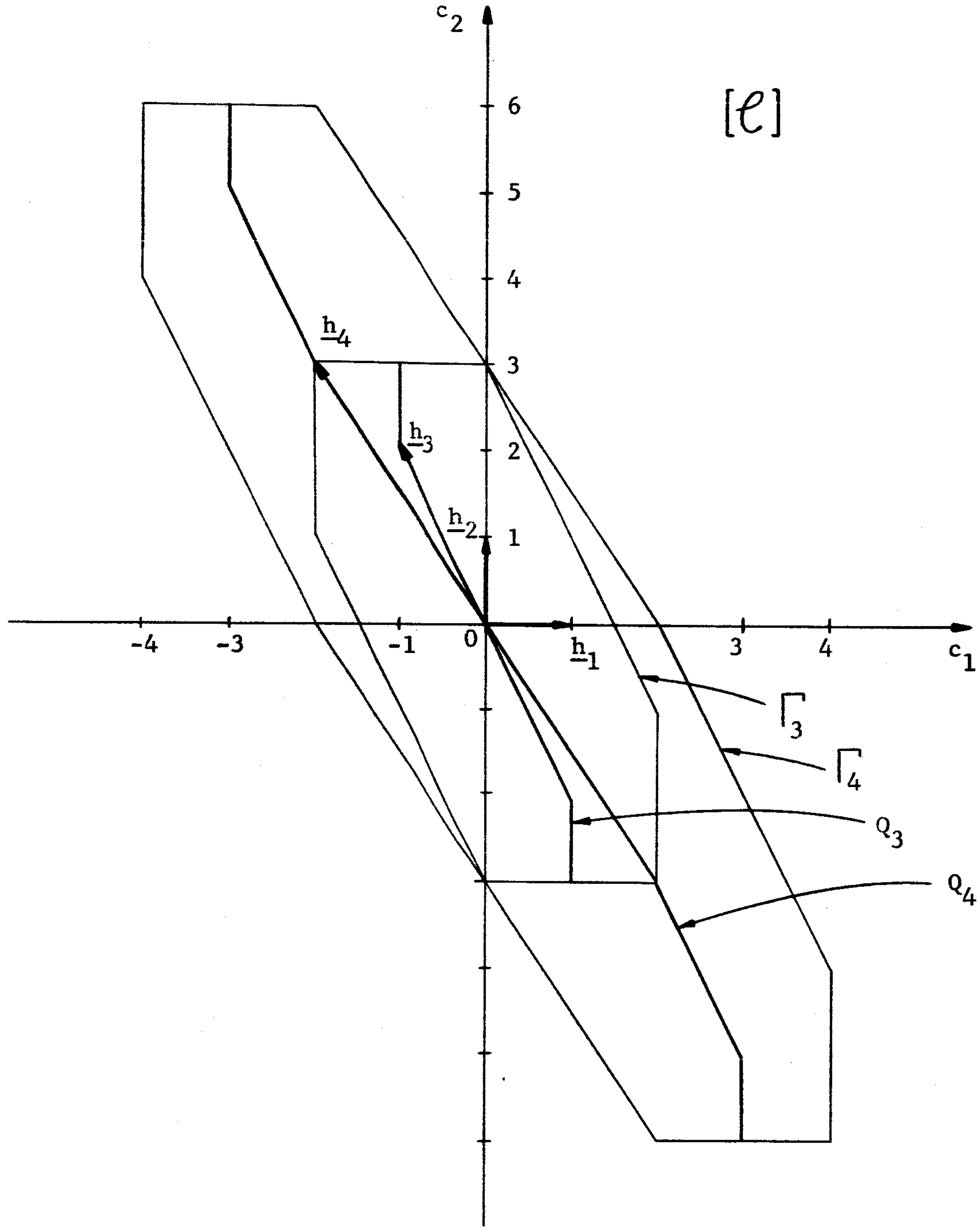


Figure 64. Alternative sets Q_3 and Q_4 for the plant $1/s^2$.

plant have the orthogonality property,

$$\underline{h}_i^t \, \underline{h}_j = 0 \quad , \tag{4-54}$$

if i is an even integer and j is an odd integer. Consider the odd numbered invariant vectors lying along the line $c_1 = 0$. These vectors cannot be utilized in any way to help represent a component c_2 of an initial state $\underline{c}$. Similarly the even numbered vectors, lying along the line $c_2 = 0$, cannot contribute towards a representation of the component c_1 of the initial state. The input sequence may therefore be divided into two parts, one part containing the odd numbered members, u(1), u(3), ..., and the other the even numbered members, u(2), u(4), ... Then, compare Equations (3-133) and (3-134), the deadbeat constraint, Equation (4-4) becomes with N arbitrarily chosen even,

$$c_1 = u(1) - u(3)e^{2aT} + u(5)e^{4aT} - \dots (-1)^{(N-2)/2}u(N-1)e^{(N-2)aT} , \tag{4-55}$$

$$c_2 = u(2) - u(4)e^{2aT} + u(6)e^{4aT} - \dots (-1)^{(N-2)/2}e^{(N-2)aT} . \tag{4-56}$$

Now compare the first order plant,

$$G_p(s) = \frac{1}{s + 2a} . \tag{4-57}$$

The initial state for this plant is a scalar quantity, let it be c, and the deadbeat constraint is

$$c = u(1) + u(2)e^{2aT} + u(3)e^{4aT} + \dots + u(N)e^{(N-1)2aT} . \tag{4-58}$$

An example of the fuel optimum input sequence for the plant of Equation

(4-57) is shown in Figure 53, page 138. Similarly the fuel optimum input sequence for a tuned second order plant is shown in Figure 65. Figure 65 is a graphical method of obtaining the sequence, and N is chosen as $N = 6$. The two parts of Figure 65 are seen to be identical except for the labelling of the input members.

The input sequence may also be obtained in a closed loop manner. Since the first input, $u^f(1)$, is always independent of the component c_2, it is only necessary to define Q_N for the component c_1. It is convenient, however, to give $u^f(1)$ in the equivalent graphical form shown in Figure 66. Since $\underline{h}_6$ does not lie on the c_1 axis of $\mathcal{C}$-space, Figure 66 is applicable to settling times of five or six sampling periods.

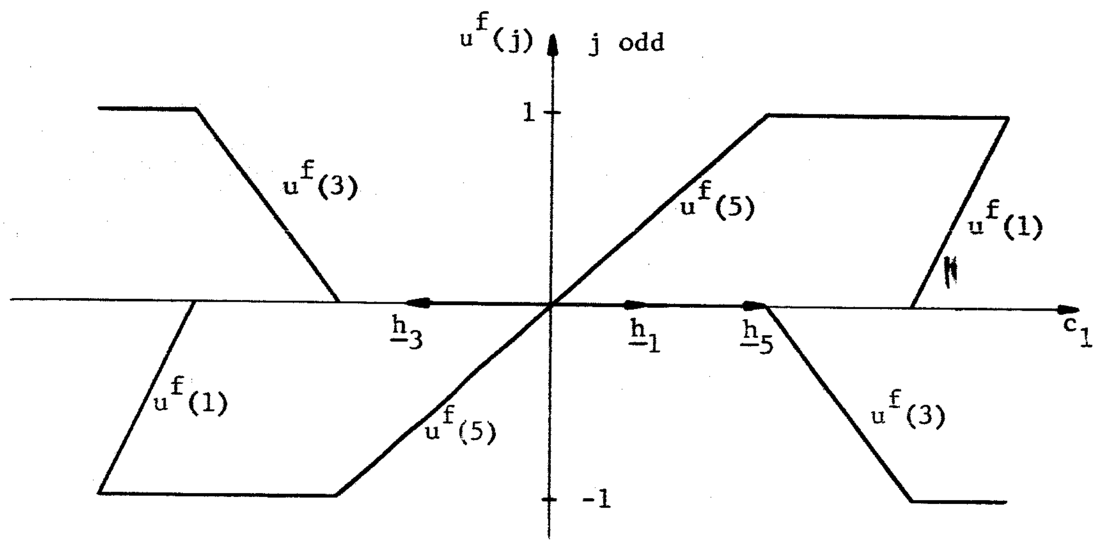
To conclude this discussion of the fuel optimum sequence for tuned plants consider Figure 67, which shows the set Γ_6 containing a given initial state $\underline{c}$. Figure 65 gives the fuel optimum representation as

$$\underline{c} = 0\ \underline{h}_1 + 0.5\ \underline{h}_2 - 0.5\ \underline{h}_3 - \underline{h}_4 + \underline{h}_5 + \underline{h}_6 \quad . \tag{4-59}$$

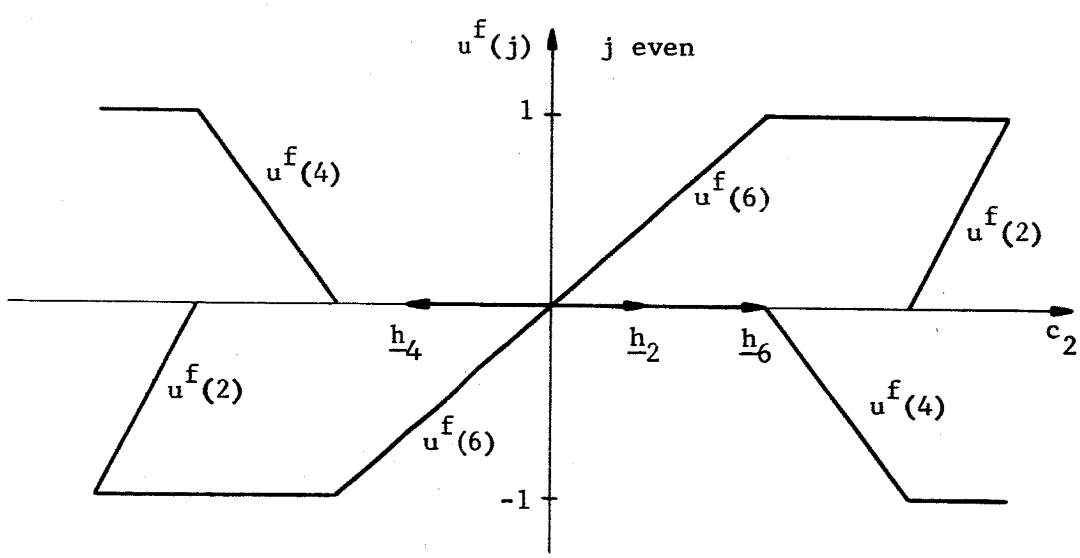
Since the first input is zero, $u^f(1) = 0$, the system is allowed to run freely for one sampling period, after which the state has reached the point $\underline{c}_1$ given by

$$\underline{c}_1 = 0.5\ \underline{h}_1 - 0.5\ \underline{h}_2 - \underline{h}_3 + \underline{h}_4 + \underline{h}_5 \quad . \tag{4-60}$$

The plant then receives an input of +0.5, and moves to the point $\underline{c}_2$, where

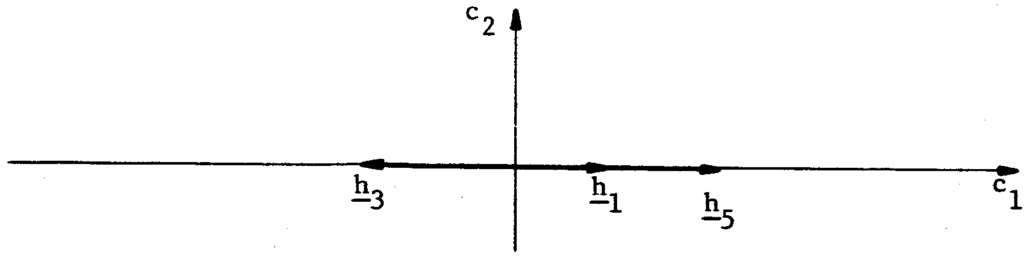


a. Odd numbered input members

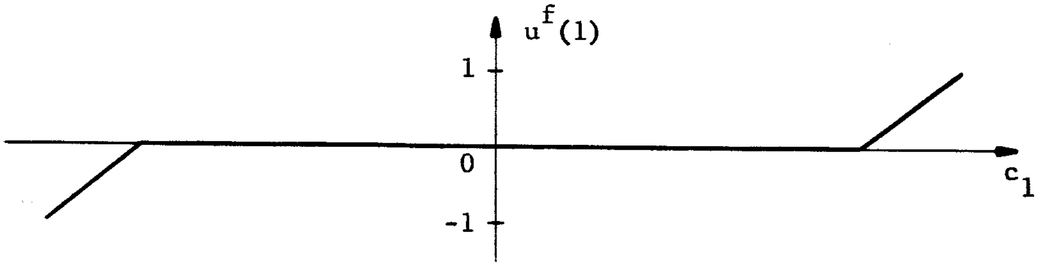


b. Even numbered input members

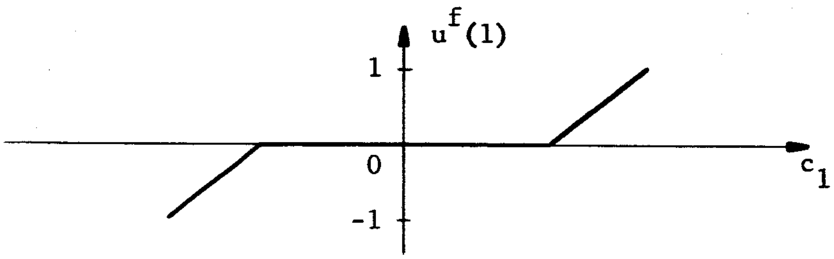
Figure 65. The fuel optimum input sequence (open loop) for a tuned second order plant with N = 6.



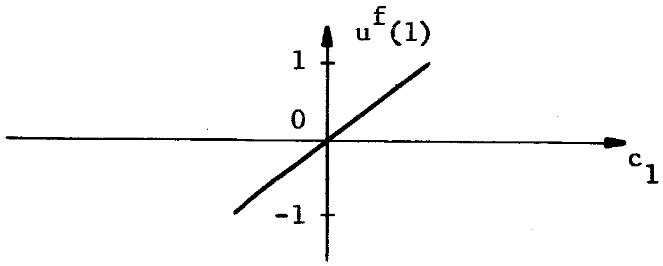
a. The invariant vectors $\underline{h}_1$, $\underline{h}_3$ and $\underline{h}_5$



b. N = 5 or 6



c. N = 3 or 4



d. N = 1 or 2

Figure 66. Graphical forms of the closed loop method for finding the fuel optimum input sequence.

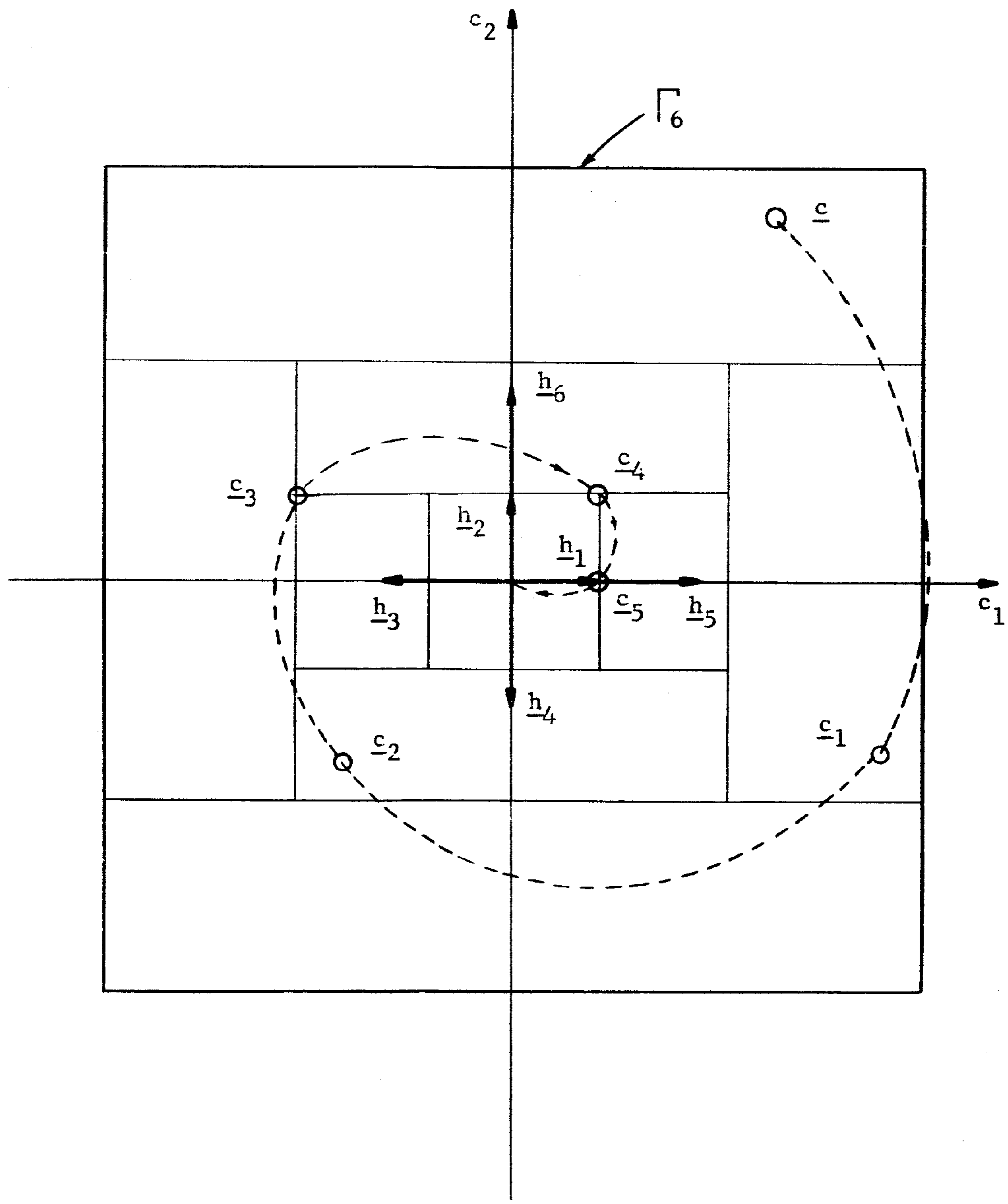


Figure 67. A fuel optimum trajectory for a tuned second order system with N = 6.

$$\underline{c}_2 = -0.5\,\underline{h}_1 - \underline{h}_2 + \underline{h}_3 + \underline{h}_4 \quad . \tag{4-61}$$

Over the next sampling period the input -0.5 is applied and at the end of this sampling period the plant state is at

$$\underline{c}_3 = -\underline{h}_1 + \underline{h}_2 + \underline{h}_3 \; , \tag{4-62}$$

following which the inputs -1, +1, +1 are applied successively. The trajectory between the sampling periods is indicated by the dashed line in Figure 67.

CHAPTER V

PRACTICAL IMPLEMENTATION OF THE OPTIMUM CONTROL SYSTEM

I. INTRODUCTION

The purpose of this chapter is to discuss how the theory of the preceeding chapters can best be utilized to generate the minimum energy and minimum fuel input sequences. Both open loop and closed loop methods are considered, the major portion of the chapter being concerned with the closed loop control of first and second order systems. The closed loop controllers required vary in complexity from simple direct feedback, to time-varying piecewise linear gains feeding a logic unit.

II. CLOSED LOOP VERSUS OPEN LOOP CONTROL

The configuration of the controlled plant and the controller is shown in Figure 2, page 3. The controller receives information on the state of the plant through identification of the state variables, $x_1(t), \ldots, x_n(t)$. Not all of these variables may be available, and some may therefore have to be estimated. However, it is tacitly assumed that, whenever necessary, the state vector,

$$\underline{x}(t) = \begin{bmatrix} x_1(t) \\ \cdot \\ \cdot \\ \cdot \\ x_n(t) \end{bmatrix} , \qquad (5\text{-}1)$$

can be found at each sampling period, $t = kT$, $k = 0, 1, \ldots$. Having been supplied with the state vector, the way in which the controller acts on this information determines whether the system operates in an open loop or closed loop manner. The desired state of the plant is at the origin of the state space and at time $t = 0$ the plant is in some disturbed state, $\underline{x}(0) \neq 0$. The controller is allowed a total time of NT seconds to bring the plant to the desired state in some optimal fashion. If the controller, being given the state $\underline{x}(0)$, generates the entire optimum input sequence, $u(1), u(2), \ldots, u(N)$, on the basis of the state $\underline{x}(0)$ alone, the control is said to be open loop. If, however, the controller is structured so that it requires knowledge of the plant state at each sampling period in order to generate the optimum input sequence, the resulting control is said to be closed loop. Open loop control has the disadvantage that if the system encounters any disturbances during the time interval $0 < t \leqslant NT$, the primary mission, that of bringing the state to the origin, will almost inevitably fail to be accomplished. On the other hand, closed loop systems, being based on a feedback principle, can still complete the primary mission if the disturbances are not too severe. It is beyond the scope of this dissertation, however, to attempt to discuss the various cited advantages and disadvantages of each method.

III. OPEN LOOP CONTROL

The main body of this dissertation has been concerned with open loop methods for solving the minimum fuel and energy problems. The open loop method of solution follows as a natural consequence of the particular approaches used on the problems. The one exception is, of course, the closed loop approach to the minimum fuel problem with input saturation discussed in the latter half of Chapter IV. The open loop approaches to the various problems are now presented. In such methods, the entire optimum input sequence is calculated, stored, and fed to the plant piece by piece.

The Linear Minimum Energy Input Sequence

For the general n-th order plant, the linear energy optimum input sequence, $\underline{u}^o$, can be calculated directly from Equation (2-37):

$$\underline{u}^o = C^t \left[CC^t\right]^{-1} \underline{x}(0) \,, \tag{5-2}$$

where C is the N x n matrix given in Equation (2-4). The inversion required to obtain the n x n matrix $\left[CC^t\right]^{-1}$ can be performed on a digital computer if, say, $n > 4$. If the sequence $\underline{u}^o$ is to be studied over a range of initial states, the plant is of the first or second order and N is not too large, then graphical methods may be more convenient. The graphical approach is given in Chapter II, and is based on Equation (2-25),

$$\underline{b}^o = H^t \, \underline{a}^o \quad , \tag{5-3}$$

relating the two parts of $\underline{u}^o$.

Both of these methods are described in detail by the example on page 58, which concerns itself with the plant $1/s^2$ with the settling time given as four sampling periods.

The Linear Minimum Fuel Input Sequence

First order systems are solved, the solution being given by Equations (2-74) through (2-77). The plant

$$G_p(s) = \frac{1}{s} \tag{5-4}$$

has an infinite number of possible input sequences. Two of the more obvious ones are given in Equations (2-75) and (2-76). Second order systems are solved if the state $\underline{c}$ can be identified as belonging to one of the cones $C_s(\pm i, \pm j)$. Equations (2-84) and (2-85) give the solution when such an identification has been made. The minimum fuel input sequence may not be unique, and this knowledge can prove useful. For example, consider the plants,

$$G_p(s) = \frac{1}{s^2} \tag{5-5}$$

and

$$G_p(s) = \frac{1}{s(s + \lambda)} \quad . \tag{5-6}$$

The set $S_5(1)$ for the plant of Equation (5-5) and an example of the set $S_4(1)$ for the plant of Equation (5-6) are shown in Figure 14, page 48 and Figure 15, page 49 respectively. These figures show by the cross-hatched areas the regions in which the minimum fuel input sequence is not unique. It will be noted that the invariant vector $\underline{h}_1$ can always be used in conjunction with $\underline{h}_N$ to give a fuel optimum input sequence for any initial state. Therefore, if the plant is given by either Equation (5-5) or (5-6), the optimum input sequence can always be obtained as

$$\begin{bmatrix} u(1) \\ \\ u(N) \end{bmatrix} = \left[\underline{h}_1 \ \underline{h}_N\right]^{-1} \underline{c} \ ; \quad u(2), \ldots, u(N-1) = 0 \ . \tag{5-7}$$

Equation (2-114) gives such an input sequence for the minimum fuel example beginning on page 64. As another example of a second order system for which the fuel optimum input sequence is readily obtained, consider the case of a plant of the form

$$G_p(s) = \frac{1}{(s + a + jb)(s + a - jb)} \tag{5-8}$$

where the tuning condition of Equation (3-130) is satisfied; i.e.,

$$bT = \frac{\pi}{2} \quad . \tag{5-9}$$

The first six invariant vectors are shown in Figure 52, page 131, for the case $a > 0$. The components c_1 and c_2 of the initial state $\underline{c}$ can therefore be represented by Equations (3-135) and (3-136):

$$c_1 = u(1) + \sum_{j=3}^{N} u(j)(-1)^{(j-1)/2} e^{(j-1)aT} \ , \quad j \text{ odd} \ , \tag{5-10}$$

$$c_2 = u(2) + \sum_{j=4}^{N} u(j)(-1)^{(j-2)/2} e^{(j-2)aT} \ , \quad j \text{ even}. \tag{5-11}$$

Without loss of generality, suppose that N is an even integer. The invariant vectors $\underline{h}_N$ and $\underline{h}_{N-1}$ are therefore given by

$$\underline{h}_N = \begin{bmatrix} 0 \\ \\ (-1)^{(N-2)/2} e^{(N-2)aT} \end{bmatrix} \tag{5-12}$$

and

$$\underline{h}_{N-1} = \begin{bmatrix} (-1)^{(N-2)/2} e^{(N-2)aT} \\ \\ 0 \end{bmatrix} . \tag{5-13}$$

If $a > 0$, these are the longest invariant vectors, and, compare Equation (2-74), the unique fuel optimum input sequence is

$$\left.\begin{aligned} u(1) &= u(2) = \ldots = u(N-2) = 0 \\ u(N-1) &= c_1(-1)^{(N-2)/2} e^{-(N-2)aT} \\ u(N) &= c_2(-1)^{(N-2)/2} e^{-(N-2)aT} \end{aligned}\right\} . \tag{5-14}$$

If $a = 0$, all the invariant vectors are of unit length, corresponding to the plant

$$G_p(s) = \frac{1}{(s^2 + b^2)} , \qquad bT = \frac{\pi}{2} , \tag{5-15}$$

and one solution to the minimum fuel problem would be Equation (5-14) with $a = 0$. Another solution, compare Equation (2-75), with N even, is

$$\begin{aligned} u(1) &= - u(3) = \ldots = (-1)^{(N-2)/2} u(N-1) = 2c_1/N \\ u(2) &= - u(4) = \ldots = (-1)^{(N-2)/2} u(N) = 2c_2/N \end{aligned} . \tag{5-16}$$

Finally, if $a < 0$, corresponding to an unstable plant of the form of Equation (5-8), the unique fuel optimum solution is, compare Equation (2-77),

$$u(1) = c_1,\ u(2) = c_2,\ u(3) = u(4) = \ldots = u(N) = 0 . \tag{5-17}$$

More general second order systems, as mentioned above, are solved by Equations (2-84) and (2-85). If the initial state $\underline{c}$ lies in the cone $C_s(\underline{+}i,\underline{+}j)$, the state can be uniquely represented as

$$\underline{c} = \mu_1 \underline{h}_i + \mu_2 \underline{h}_j \quad , \tag{5-18}$$

giving the minimum fuel input sequence as

$$u(i) = \mu_1,\ u(j) = \mu_2,\ u(k) = 0,\ k = 1, 2, \ldots, N,\ k \neq i,j. \tag{5-19}$$

If N is large, and the plant has real poles, the state space may be partitioned by many cones, and, even with the help of a digital computer, it may be a problem to identify the state as belonging to a particular cone. As is shown later, a closed loop procedure may help to solve this problem, but hybrid techniques are necessary.

Higher order systems can, in principle, be treated by the techniques underlying Equations (5-18) and (5-19), but in practice, since the cones must be defined and identified in n-dimensions, it may be a very difficult task to obtain the solution. A general method would be to use a linear programming technique.

The Minimum Energy Input Sequence with Saturation

The problems involved in obtaining the optimum input sequence when the input members are subject to amplitude constraints are discussed in detail in Chapter III. If the linear minimum energy input sequence, $\underline{u}^o$, has members which exceed the saturation limits, it is shown that the constrained minimum energy solution, $\underline{u}^e$, must have one or more of its members equal to the saturation limit. Theorems 2 and 3, given on pages 103 and 120 respectively, can be used to find which members of $\underline{u}^e$ are equal to the limit.

If only one member of $\underline{u}^o$ exceeds the saturation limit,

$$\left|u^o(j)\right| > 1, \quad \left|u^o(i)\right| < 1, \quad i = 1, 2, \ldots, N, \quad i \neq j, \tag{5-20}$$

Theorem 2 says that

$$u^e(j) = \text{sgn. } u^o(j) \quad . \tag{5-21}$$

The problem then becomes: minimize

$$\sum_{\substack{i=1 \\ i\neq j}}^{N} u(i)^2 \tag{5-22}$$

subject to

$$\underline{c} - \left[\text{sgn. } u^o(j)\right] \underline{h}_j = \sum_{\substack{i=1 \\ i\neq j}}^{N} u(i)\, \underline{h}_i \; ; \quad \left|u(i)\right| \leqslant 1 \; . \tag{5-23}$$

If the solution to this second problem has all its members lying within the saturation limits, the original is solved. If only one of its members exceeds the saturation limit, Theorem 2 again guarantees that Equation (5-21) gives the corresponding member of $\underline{u}^e$. Suppose that, on continuing in this manner, Theorem 2 is applicable for each new problem; i.e., no more than one member of each corresponding linear minimum energy input sequence exceeds the saturation limits: eventually, assuming of course that the initial state $\underline{c}$ is in Γ_N, there will result a problem whose linear energy optimum input sequence satisfies the saturation contraint. This solution, combined with all the members which were given by Equation (5-21), constitutes the sequence $\underline{u}^e$.

If more than one member of $\underline{u}^o$ exceeds the saturation limit, Theorem 3 is applicable. Theorem 3 gives two conditions which must be satisfied before Equation (5-21) can be used to give the members of $\underline{u}^e$. The first condition is Equation (3-105), and this may be verified by a simple computation. The other condition which must be satisfied is stated in Equations (3-106) and (3-107). In general, it would be a very complicated task to check this condition each time Theorem 3 was applicable. Therefore, since it seems likely that this condition will rarely be violated, it is suggested that when more than one member of a linear minimum energy sequence exceeds the saturation limit, only the test of Equation (3-105) be used to determine for which of these members Equation (5-21) is applicable. If, on following the step by step procedure outlined above, there eventually results a linear energy optimum sequence which does satisfy the saturation constraints, the omission of the second condition will have been justified. On the other hand, if it eventually becomes obvious that it is now impossible to take the state into the origin with the constrained inputs associated with the remaining invariant vectors, then Equation (5-21) has been applied incorrectly to one or more of the input members. An example of this is given on page 110.

It was shown in Chapter III that first order plants and second order plants with integration or tuned complex poles can always be solved by the systematic use of Equation (5-21). The examples on page 76 and page 129 give the minimum energy input sequence for a first order plant and the plant $1/s^2$ respectively.

The open loop technique of nonlinear programming is a general method which can always be used to obtain a solution to the minimum energy problem with input saturation (31, 32).

The Minimum Fuel Input Sequence with Saturation

Chapter IV discusses the general problem of obtaining the minimum fuel input sequence with input saturation. The optimum sequence for first order plants and second order underdamped plants with tuning can be obtained quite easily in open loop form. Figure 53, page 138, gives an example of the optimum sequence in graphical form for a first order plant. Figure 65, page 168, gives the optimum sequence for the tuned plant in a graphical form. Second order plants with real poles and untuned underdamped plants are best treated by closed loop techniques. Alternatively, they can be approached in an open loop manner by the use of linear programming (29, 31). Linear programming may also be used as a general method of obtaining the amplitude constrained fuel optimum input sequence for higher order systems.

IV. CLOSED LOOP CONTROL

If the system has closed loop control, it is generally implied that the input sequence is generated as

$$u(t) = f\left[x_1(t),\ x_2(t),\ \ldots,\ x_n(t),\ t\right] \tag{5-24}$$

where f is some scalar function of the state vector $\underline{x}(t)$ and the time t. The most general form of Equation (5-24) that is required to cover the cases discussed below is

$$u(k + 1) = f\left[\underline{x}(k),\ kT\right], \quad k = 0, 1, \ldots, N-1 \quad . \tag{5-25}$$

Equation (5-25) means that the control level, u(k), over the time interval $(k - 1)T < t \leqslant kT$ is obtained from some function of the state variables at the time $t = (k - 1)T$, and, as implied, this function may not be the same at each sampling instant.

The Form of the Feedback Function

Equation (5-25) may take several different forms. Before discussing the closed form solutions to the minimum energy and fuel problems, it is useful to classify some of the different types of feedback that will be of interest.

Time-invariant (constant) linear feedback. Figure 68 shows the controller configuration. The controller gives the input sequence as

$$u(k) = f_1\, x_1(k) + f_2\, x_2(k) + \ldots + f_n\, x_n(k) \tag{5-26}$$

where $f_1, f_2, \ldots, f_n$ are constant. Such feedback has been used to implement linear time optimum control (7), when the coefficients f_1, $f_2, \ldots, f_n$ constitute the first row of the matrix R^{-1}. Another example where time invariant linear feedback can be used to implement an optimum control sequence is when the cost function is of the form

$$\sum_{k=1}^{N} \underline{x}(k)^t P\, \underline{x}(k) + \underline{u}^t\, S\, \underline{u} \quad , \tag{5-27}$$

where, in general, S is an n x n positive definite matrix and P is an N x N positive definite matrix. If $N \rightarrow \infty$, the optimum feedback approaches the form of Equation (5-26), (1, page 486; 17, page 1823).

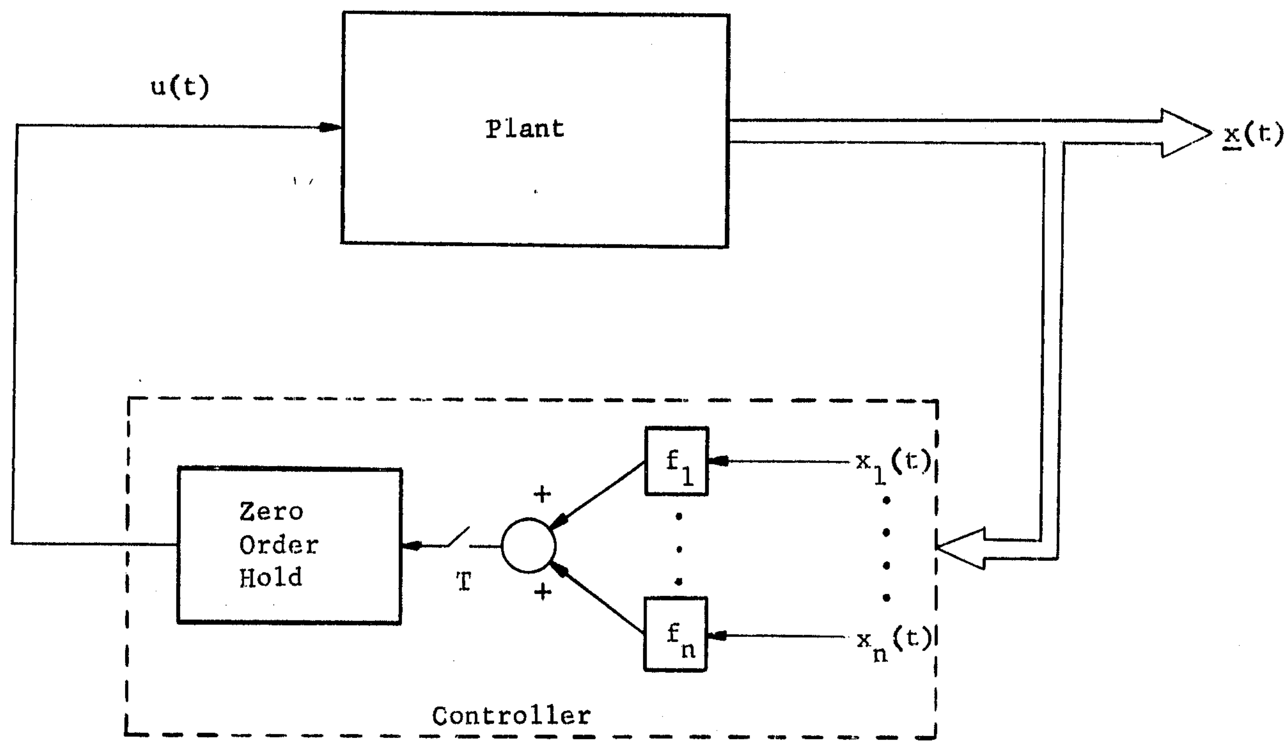


Figure 68. The system configuration with time-invariant linear feedback.

Time-invariant piecewise linear feedback. The input sequence for the time optimum deadbeat regulator with input saturation has been implemented, for second order systems with real poles, by using only a piecewise linear function of the state (14). Figure 69 shows a controller configuration of this type. Third order systems with real poles require slightly more complicated considerations (24). The linear transformation R^{-1} enables the controller to work with the state in $\mathcal{C}$ -space rather than $\mathcal{X}$ -space. The function f is a piecewise linear function of the variable $c_2(t)$. This function can be implemented with the use of analog devices (14). The output of the summing junction, $f(c_2) + c_1$, as will be shown later, represents the distance of the state $\underline{c}$, in the direction of the c_1 axis, from a line in two-dimensional $\mathcal{C}$ -space. If the state lies to the right of the line, the quantity $f(c_2) + c_1$ is positive, and if to the left of the line, $f(c_2) + c_1$ is negative. The ideal saturation function following the summation has the output sat. $\left[f(c_2) + c_1\right]$, where the sat. function is defined by Equation (4-38).

Time-varying linear feedback. If the controller employs time-varying linear feedback, the input sequence is generated as

$$u(k) = f_1(k)\ x_1(k) + f_2(k)\ x_2(k) + \dots + f_n(k)\ x_n(k) \tag{5-28}$$

$$= \underline{f}(k)\ x(k), \quad k = 1, 2, \dots, N \ . \tag{5-29}$$

The physical configuration of the controller is the same as that shown in Figure 68, except that the gains $f_1, \dots, f_n$ change, so that at each

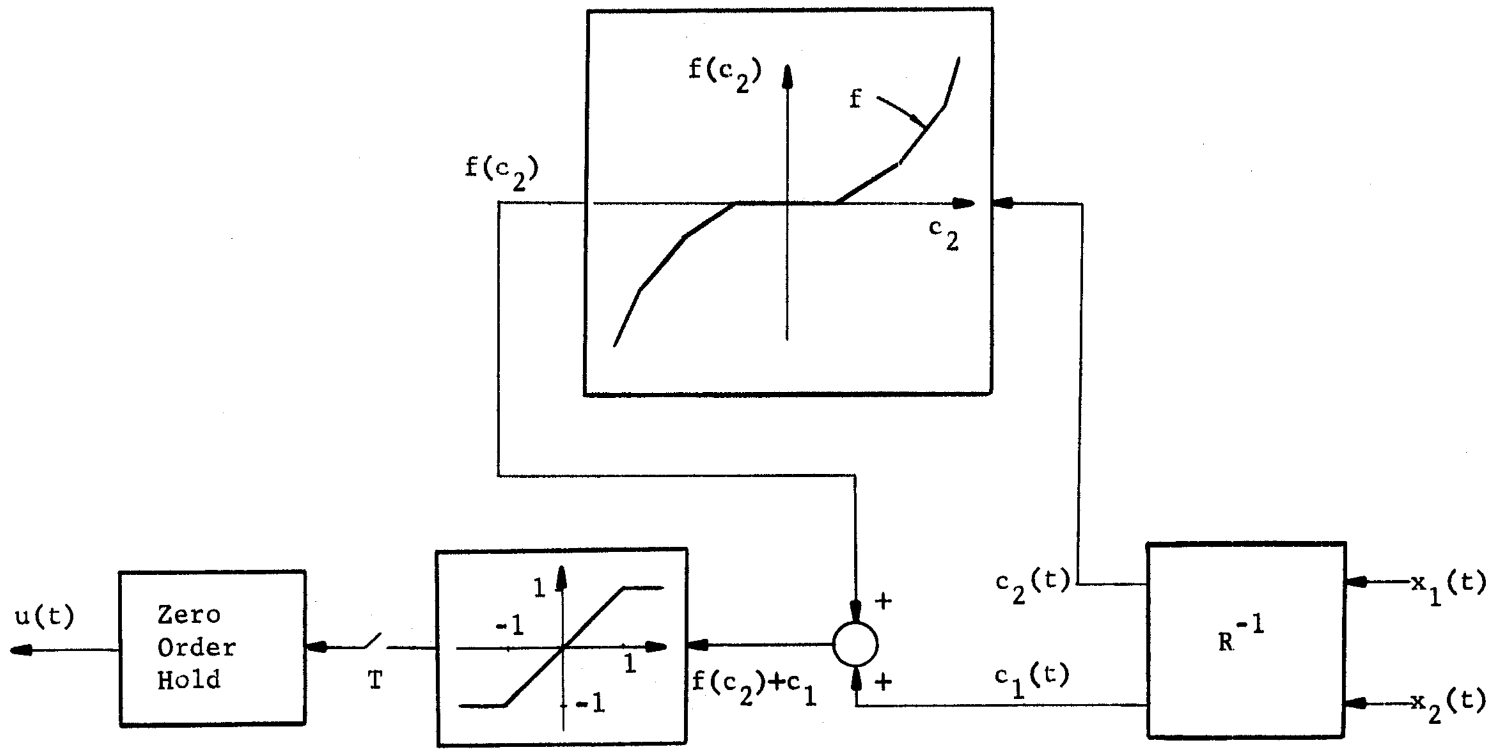


Figure 69. The controller configuration for time-invariant piecewise linear feedback for a second order system.

sampling period they have a predetermined optimum value. The cost function of Equation (5-27) requires such a feedback function if N is finite (1, 5, 17).

Time-varying piecewise linear feedback. Figure 69 shows an example of time-invariant piecewise linear feedback. Suppose the function f does not remain constant, but instead takes on different forms at each sampling instant. The resulting feedback is called time-varying piecewise linear feedback. A slightly more complex form of this type of feedback uses two such time-varying functions, each fed with the variable c_2. With the help of some elementary logic, the controller is capable of finding the distance, in the direction of the c_1 axis, of a state $\underline{c}$ from some time-varying polygonal region in a two-dimensional $\mathcal{C}$-space.

Having considered the types of feedback that may be used, the closed loop control of the minimum fuel and energy systems are now considered in detail.

V. CLOSED LOOP CONTROL FOR THE LINEAR MINIMUM ENERGY SYSTEM

The open loop solution to the linear minimum energy problem can be obtained from Equation (5-2). If the suggested settling time is N-sampling periods, the n x N matrix C in Equation (5-2) is given by Equation (2-4) as

$$C = \left[\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N\right] , \tag{5-30}$$

where

$$\underline{r}_j = - G(-jT)\ \underline{h}(T), \quad j = 1, 2, \ldots, N \tag{5-31}$$

are the first N canonical vectors, as given in Appendix A, Equation (A-19). The first member of the N-member input sequence, $\underline{u}^o$, is given by

$$u^o(1) = \underline{f}(1)\ \underline{x}(0) \quad , \tag{5-32}$$

where $\underline{f}(1)$ is the first row of the matrix $C^t\left[CC^t\right]^{-1}$, and $\underline{x}(0)$ is the initial state, given at time t = 0. Rather than calculating the second row of this matrix to obtain $u^o(2)$, the closed loop procedure requires the calculation of the first row, $\underline{f}(2)$, of the new matrix $C^t\left[CC^t\right]^{-1}$, where the matrix C is now n x N-1, and is given by

$$C = \left[\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_{N-1}\right] \quad . \tag{5-33}$$

After one sampling period the plant will have the state $\underline{x}(1)$, and the second input to be applied is then

$$u^o(2) = \underline{f}(2)\ \underline{x}(1) \quad . \tag{5-34}$$

If no disturbances were present over the first sampling period, $u^o(2)$ as given by Equation (5-34) will be exactly the same as the member that could have been obtained from the initial state, $\underline{x}(0)$, using the second row of the original N x n matrix $C^t\left[CC^t\right]^{-1}$, C being given by Equation (5-30). To continue this feedback generation of the optimum input sequence, let $\underline{f}(j + 1)$ be the first row of the matrix $C^t\left[CC^t\right]^{-1}$ when

$$C = \left[\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_{N-j}\right] , \quad j = 0, 1, \ldots, N-n \quad . \tag{5-35}$$

For $j = N-n$, the matrix C is given by

$$C = \left[\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_n\right] \tag{5-36}$$

and this may be inverted. Equation (5-36) defines the matrix R, see Equation (2-8), so that $f(N-n + 1)$ is the first row of R^{-1}. Since the object of the regulating system is to force the state to the origin, the feedback must be kept constant for the remainder of the regulation process. If no disturbances occur over the last n sampling periods, the regulation will be completed in a total of N sampling periods. If such disturbances do occur, the feedback will keep on trying to force the state into the origin.

Thus, $\underline{f}(1)$ is the first set of gains in Equation (5-29), $\underline{f}(2)$ the second set, and so on. The optimum controller uses these time-varying gains, and the configuration is shown in Figure 68, page 183. At the beginning of the regulation process the controller has the vector gain $\underline{f}(1)$, and, operating on the state $\underline{x}(0)$, gives the hold device the first optimum input level to be applied to the plant. During the time $0 < t \leqslant T$, this gain is replaced by $\underline{f}(2)$, which, at $t = T$, operates on the state $\underline{x}(1)$ to give the second input level. This process continues until only n sampling periods remain. The feedback is then kept constant at $f(N-n + 1)$, producing the last n input levels, $u^o(N-n + 1)$, $u^o(N-n + 2)$, ..., $u^o(N-1)$, $u^o(N)$.

The implementation of these time varying gains may prove too costly in practice, and it has been suggested that an approximation to the minimum energy input sequence could be obtained by using the fixed gain $\underline{f}(1)$ for the entire input sequence (48). For plants with real poles or heavily damped complex poles, a near deadbeat response is attained, and the energy consumption is reasonably small. The choice of N naturally has considerable effect on the settling time, and may therefore be used as a design parameter.

VI. CLOSED LOOP CONTROL FOR THE LINEAR MINIMUM FUEL SYSTEM

First Order Systems

Consider the first order plant,

$$G_p(s) = \frac{1}{s + \lambda} \quad . \tag{5-37}$$

When $\lambda > 0$, Equation (2-74) gives the unique fuel optimum solution as

$$u(1) = u(2) = \dots = u(N-1) = 0, \quad u(N) = c(0)/e^{(N-1)\lambda T}, \tag{5-38}$$

where c(0), a scalar, is the initial state in $\mathcal{C}$-space. The controller, having been allowed N sampling periods to bring the state to the origin, therefore waits for N-1 sampling periods, and then applies the input

$$u = c(N-1) \quad . \tag{5-39}$$

If $\lambda = 0$, there are many possible input sequences which take c(0) to the origin with minimum fuel. The input sequence of Equation (2-75),

$$u(1) = u(2) = \dots = u(N) = c(0)/N \quad , \tag{5-40}$$

minimizes not only the fuel, but also the energy. If the input sequence of Equation (2-76),

$$u(1) = c(0), \quad u(2) = u(3) = \dots = 0 \quad , \tag{5-41}$$

is chosen, the minimum fuel regulation may be accomplished in only one sampling period. If $\lambda < 0$, the input sequence of Equation (2-77) is optimum, giving the input sequence as in Equation (5-41).

The implementation of these sequences as closed loop controllers is straightforward. Consider Equation (5-39). The controller waits $(N-1)T$ seconds and then switches the state of the plant directly into the zero order hold. The implementation of the sequence of Equation (5-40) requires a time-varying gain, so that

$$u(k + 1) = c(k)/N - k, \quad k = 0, 1, 2, \ldots, N-1. \tag{5-42}$$

The sequence of Equation (5-41) requires only that the state be fed directly into the sample-hold device. Figure 70 shows how these three controllers might be implemented.

Second Order Systems

General second order systems. The principle of the closed loop procedure is as follows. The sets $\partial S_k(1)$, $k = 3, 4, \ldots, N$ are constructed as the convex hull, see page 36, of the set of 2k points,

$$\pm \underline{h}_1, \quad \pm \underline{h}_2, \quad \pm \underline{h}_3, \quad \ldots, \quad \pm \underline{h}_k \quad . \tag{5-44}$$

Suppose N-k sampling periods have elapsed since the time t = 0 when the regulation began, and the state of the plant is $\underline{c}(N-k)$. The optimum input $u(N-k + 1)$ is then found by considering $\partial S_k(1)$. If $\underline{h}_1$ does not lie on $\partial S_k(1)$, the input $u(N-k + 1) = 0$. If $\underline{h}_1$ does lie on $\partial S_k(1)$, the input $u(N-k + 1)$ may or may not be zero, depending on the location of $\underline{c}(N-k)$. Consider Figure 71, which shows the two-dimensional $\mathcal{C}$-space divided into six regions. Regions A and A^- are the cones $C_s(1,-j)$ and

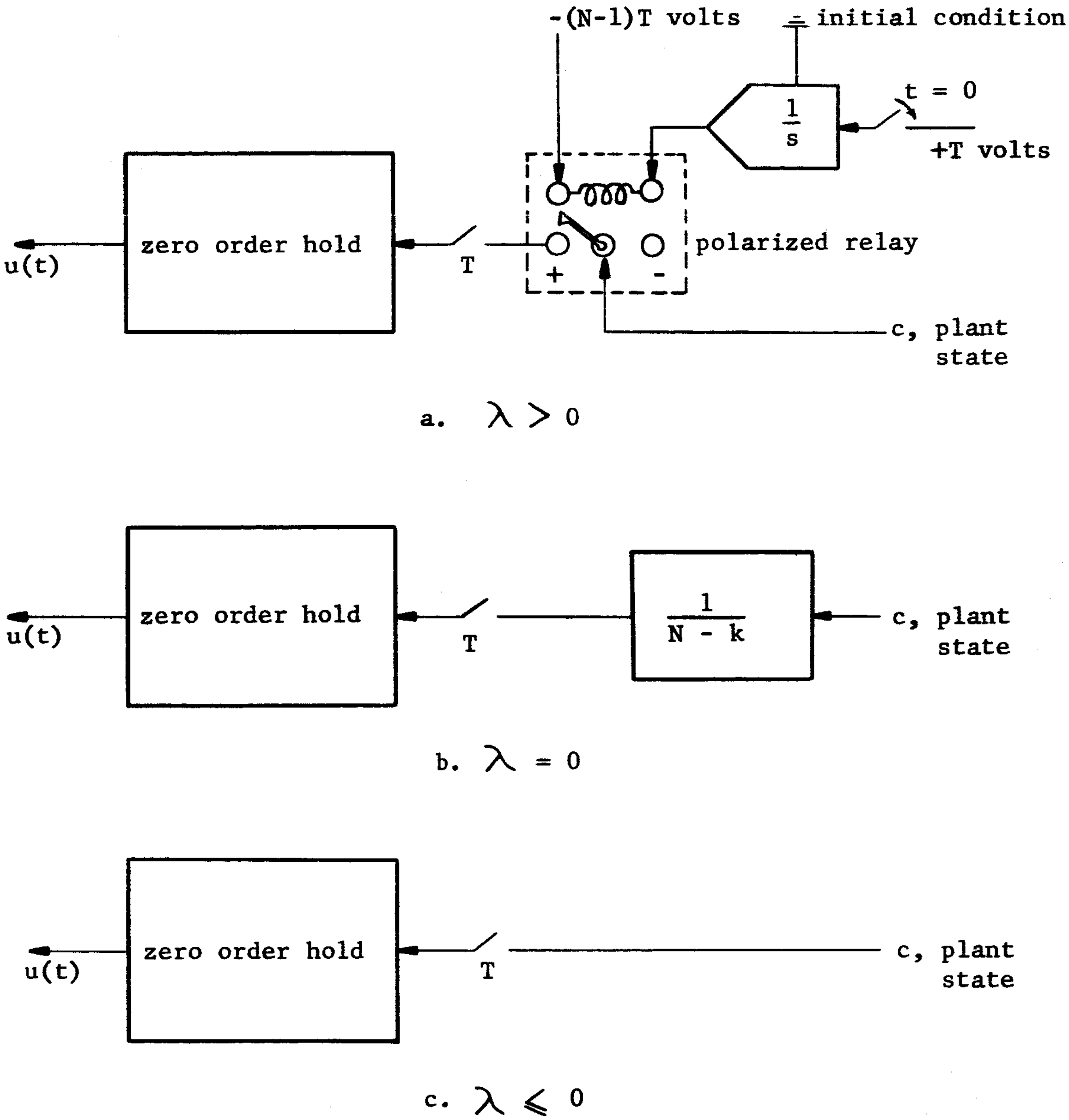


Figure 70. Closed loop implementation of the linear minimum fuel input sequence for first order systems, $1/s + \lambda$.

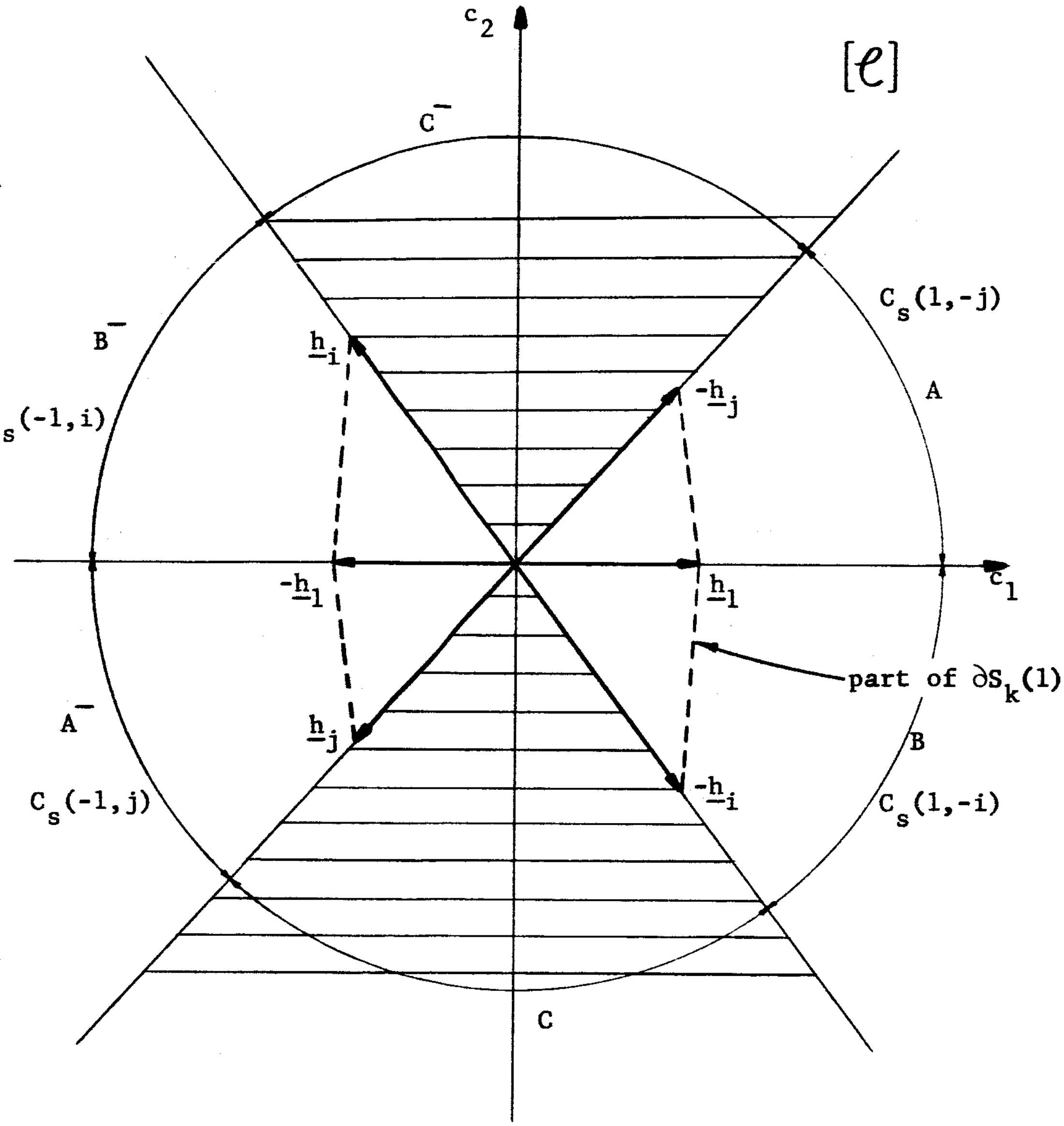


Figure 71. The invariant vector $\underline{h}_1$ lying on $\partial S_k(1)$ for a typical second order plant.

$C_s(-1,j)$. The regions B and B^- are the cones $C_s(1,-i)$ and $C_s(-1,j)$, and the cross-hatched regions, C and C^-, comprise the remainder of $\mathcal{C}$-space. If $\underline{c}(N-k)$ lies in either C or C^-, $u(N-k+1) = 0$. If $\underline{c}(N-k)$ lies in, A or A^-, $u(N-k+1)$ is given by, see Equations (2-84) and (2-85),

$$\begin{bmatrix} u(N-k+1) \\ \\ u(j) \end{bmatrix} = \begin{bmatrix} \underline{h}_1 & \underline{h}_j \end{bmatrix}^{-1} \underline{c}(N-k) \quad . \tag{5-45}$$

If $\underline{c}(N-k)$ lies in B or B^-, $u(N-k+1)$ is given by

$$\begin{bmatrix} u(N-k+1) \\ \\ u(i) \end{bmatrix} = \begin{bmatrix} \underline{h}_1 & \underline{h}_i \end{bmatrix}^{-1} \underline{c}(N-k) \quad . \tag{5-46}$$

The procedure is initiated with $k = N$. By considering $\partial S_N(1)$ the input $u(1)$ is generated. Then the set $\partial S_{N-1}(1)$ is used to give $u(2)$, and so on, until only two sampling periods remain. The remaining two input members, $u(N-1)$ and $u(N)$ are then given uniquely as

$$u(N-1) = c_1(N-2) \quad , \tag{5-47}$$

$$u(N) = c_1(N-1) \quad , \tag{5-48}$$

where $c_1(k)$ is the first component of the state $\underline{c}(k)$.

The actual implementation of this procedure by a closed loop controller is now discussed. Consider each of the regions where $u(N-k+1)$ is not zero. In region A the minimum fuel representation of the state $\underline{c}(N-k)$ is

$$\underline{c}(N-k) = u(N-k+1)\,\underline{h}_1 - u(j)\,\underline{h}_j \quad . \tag{5-49}$$

In region A^-, the representation is

$$\underline{c}(N-k) = - u(N-k + 1)\ \underline{h}_1 + u(j)\ \underline{h}_j \quad . \tag{5-50}$$

In regions B and B^-, the representations are, respectively,

$$\underline{c}(N-k) = u(N-k + 1)\ \underline{h}_1 - u(i)\ \underline{h}_i \tag{5-51}$$

$$\underline{c}(N-k) = -u(N-k + 1)\ \underline{h}_1 + u(i)\ \underline{h}_i \quad . \tag{5-52}$$

In any one of these four regions, $u(N-k + 1)$ is simply the distance of the state $\underline{c}(N-k)$, in the direction of $\pm\ \underline{h}_1$, from the cross-hatched region bounded by the lines $\mu\ \underline{h}_j$ and $\mu\ \underline{h}_i$, $-\infty < \mu < \infty$. The sign of $u(N-k + 1)$ is positive if $\underline{c}(N-k)$ lies in regions A or B, and negative if it lies in A^- or B^-. Let

$$\underline{h}_i = \begin{bmatrix} h_{i1} \\ \\ h_{i2} \end{bmatrix} \ , \quad \underline{h}_j = \begin{bmatrix} h_{j1} \\ \\ h_{j2} \end{bmatrix}, \quad j = 2, 3, \ldots, N, \tag{5-53}$$

and define

$$f_i(c_2) = - \frac{h_{i1}}{h_{i2}}\ c_2 \quad , \tag{5-54}$$

$$f_j(c_2) = - \frac{h_{j1}}{h_{j2}}\ c_2 \quad . \tag{5-55}$$

The quantity $f_j(c_2^*)$, for example, is the horizontal distance between the c_2 axis, at the point c_2^* and the line $\mu\ \underline{h}_j$, $-\infty < \mu < \infty$. Suppose the state $\underline{c}(N-k)$ lies in either region A or A^-. It can be seen that

$$u(N-k - 1) = a = f_j\Big[c_2(N-k)\Big] + c_1(N-k) \quad . \tag{5-56}$$

Similarly, if $\underline{c}(N-k)$ is in region B or B^-,

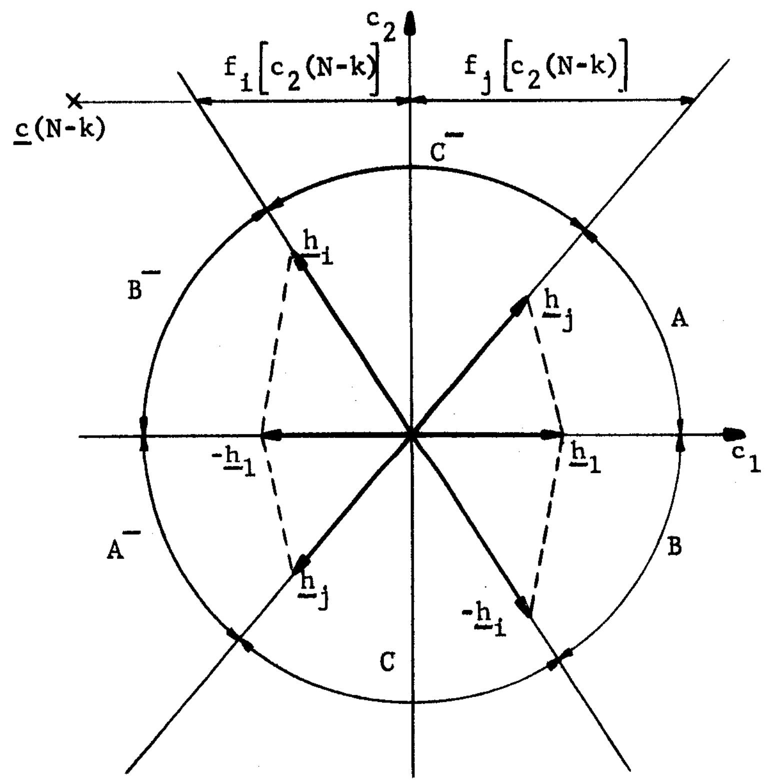
$$u(N-k-1) = b = f_i\left[c_2(N-k)\right] + c_1(N-k) \quad . \tag{5-57}$$

Figure 72 shows how the value of a from Equation (5-56), b from Equation (5-57), and $c_2(N-k)$ can be combined in a logical manner to produce the optimum value of $u(N-k+1)$. For example, suppose both a and b are positive and $c_2(N-k)$ is negative. The correct value of $u(N-k+1)$ is therefore $a = f_i\left[c_2(N-k)\right] + c_1(N-k)$. Figure 73 shows the structure of the optimum controller. The gains $f_i(c_2)$ and $f_j(c_2)$ are found from the sets $\partial S_N(1), \ldots, \partial S_3(1)$. If for any k, $\underline{h}_1$ lies interior to $\partial S_k(1)$, the f_i and f_j are chosen to give a and b of opposite sign so that the resulting input is zero. For the last two sampling periods the gains f_i and f_j are to have zero slope, so that u(N-1) and u(N) are given by Equations (5-47) and (5-48). The logic remains unchanged throughout the regulation process.

Second order systems with integration. Equation (5-5) and (5-6) describe second order systems with integration. Suppose that N-k sampling periods have elapsed since the regulation was started at time t = 0, and that the state of the plant has moved from $\underline{c}(0)$ to $\underline{c}(N-k)$. A linear fuel optimum input over the next sampling period can be found from Equation (5-7) as

$$\begin{bmatrix} u(N-k+1) \\ \\ u(k) \end{bmatrix} = \left[\underline{h}_1 \; \underline{h}_k\right]^{-1} \underline{c}(N-k) \quad . \tag{5-58}$$

Defining the feedback as



a. The regions defining u (N-k + 1)

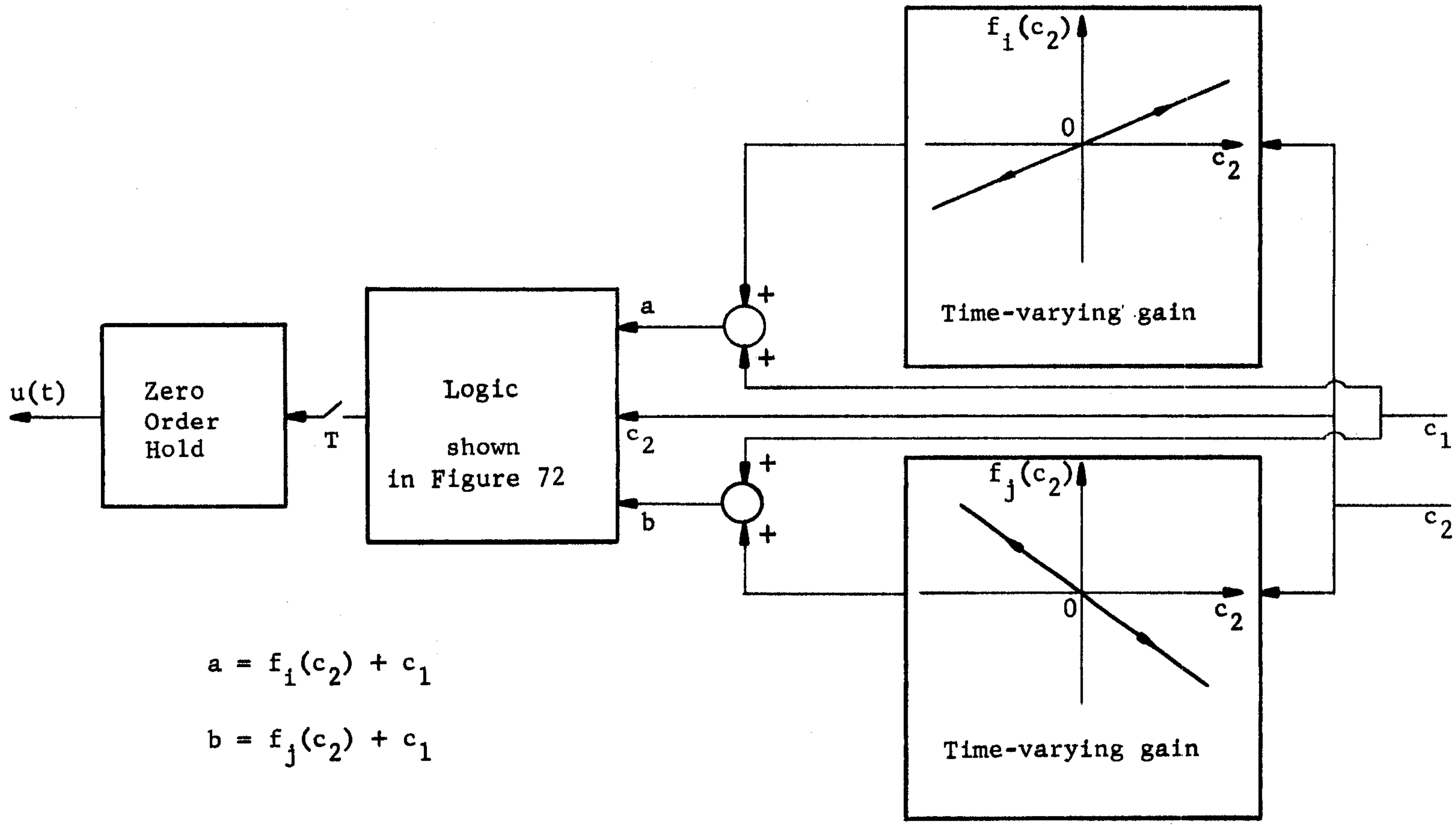
$$a = f_i\left[c_2(N\text{-}k)\right] + c_1(N\text{-}k)$$

$$b = f_j\left[c_2(N\text{-}k)\right] + c_2(N\text{-}k)$$

Region	a	b	c_2(N-k)	Input Member u (N-k + 1)
A	+	+	+	b
A^-	-	-	-	b
B	+	+	-	a
B^-	-	-	+	a
C	-	+		0
C^-	+	-		0

b. The logic used to obtain u (N-k + 1)

Figure 72. The method of obtaining u (N-k + 1) when $\underline{h}_1$ lies on $\partial S_k(1)$.



$$a = f_i(c_2) + c_1$$

$$b = f_j(c_2) + c_1$$

Figure 73. A closed loop controller giving the linear minimum fuel input sequence for a second order plant.

$$f_k(c_2) = -\frac{h_{k1}}{h_{k2}} c_2 , \quad k = 3, 4, \ldots, N, \tag{5-59}$$

gives the desired optimum input to the zero order hold as $f_k(c_2) + c_1$. Over the last two sampling periods, u(N-1) and u(N) are again given by Equations (5-47) and (5-48), so that the last two gains are $f_2 = 0$, $f_1 = 0$. The linear fuel optimum controller configuration, shown in Figure 74, is much simpler when the plant has integration.

Second order systems with tuned complex poles. Figure 52, page 131, shows the invariant vectors for a plant of the form

$$G_p(s) = \frac{1}{(s + a + jb)(s + a - jb)} \quad , \quad a > 0 \quad , \tag{5-60}$$

when the tuning condition, Equation (5-9), is satisfied. If $a > 0$, see Equation (5-14), all the input members are zero, except the last two. In closed loop form, the controller feeds nothing back until $t = (N-2)T$. The first component of the state in $\mathcal{C}$ -space is then fed directly into the zero order hold. The controller configuration is very similar to that shown in Figure 70, page 191, for the case $\lambda > 0$. The optimum controller for the case $a \leq 0$ is also directly comparable to the corresponding case $\lambda \leq 0$ shown in Figure 70.

VII. CLOSED LOOP CONTROL FOR MINIMUM ENERGY WITH INPUT SATURATION

First Order Systems

The first order plant is given by

$$G_p(s) = \frac{1}{s + \lambda} \quad . \tag{5-61}$$

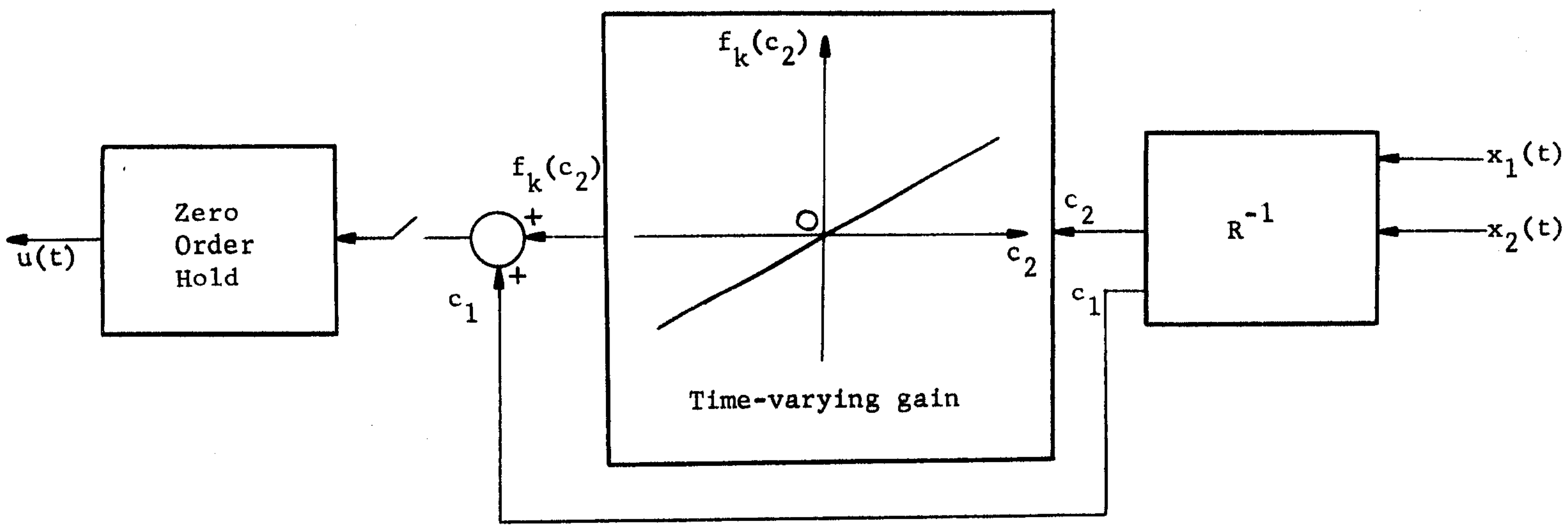


Figure 74. The closed loop controller giving a linear minimum fuel input sequence for second order plants with integration.

It is assumed that the initial state, c(0), a scalar, lies in the set Γ_N, so that, from Equation (3-8),

$$|c(0)| \leqslant \sum_{j=0}^{N-1} e^{j\gamma} \quad , \tag{5-62}$$

where $\gamma = \lambda T$. Assume further that only k of the original N sampling periods remain to complete the regulation and that the state of the plant has moved from c(0) to c(N-k). The remaining members of the linear open loop energy optimum input sequence are then given, from Equation (3-3), as

$$u^o(N-k+1+j) = e^{j\gamma} \; c(N-k) \sum_{i=0}^{k-1} e^{2i\gamma}, \quad j = 0, 1, \ldots, k-1. \tag{5-63}$$

Now, from Equation (5-62), the state c(N-k) may be assumed to lie anywhere in the range

$$-\sum_{i=0}^{k-1} e^{i\gamma} \leqslant c(N-k) \leqslant \sum_{i=0}^{k-1} e^{j\gamma} \quad . \tag{5-64}$$

Assume, without loss of generality, that c(N-k) > 0, and consider how $u^e(N-k+1)$ varies as c(N-k) moves from the origin to its extreme positive value. There are three possible cases: $\lambda > 0$, $\lambda = 0$ and $\lambda < 0$.

<u>The case $\lambda > 0$</u>. Figure 26, page 70, shows that as the state moves from the origin in the positive direction, there will come a point where the last input, $u^o(N)$, is equal to the saturation limit, +1. From

Equation (5-63) this point is given by

$$c(N-k) = e^{-(k-1)\gamma} \sum_{i=0}^{k-1} \epsilon^{2i\gamma} \quad . \tag{5-65}$$

For future convenience, define

$$d(k - 1) = \sum_{i=0}^{k-1} e^{2i\gamma} = \frac{e^{2k\gamma} - 1}{e^{2\gamma} - 1} \quad . \tag{5-66}$$

Equation (3-10) says that if

$$c(N-k) \geqslant e^{-(k-1)\gamma} d(k - 1) \quad , \tag{5-67}$$

then

$$u^e(N) = 1 \quad . \tag{5-68}$$

As $c(N-k)$ increases up to the point $e^{-(k-1)\gamma} d(k - 1)$, $u^o(N-k + 1)$ also increases, and, from Equation (5-63) with $j = 0$, reaches the value

$$u^e(N-k + 1) = e^{-(k-1)\gamma} \quad . \tag{5-69}$$

If $c(N-k)$ passes the point where $u^o(N) = 1$, Equation (5-68) gives $u^e(N) = 1$. Since only k sampling periods remain, the invariant vector associated with $u(N)$ is the k-th invariant vector, which has the length $e^{(k-1)\gamma}$. Therefore, in order to continue, Equation (5-63) must be modified to

$$u^o(N-k + 1 + j) = \frac{e^{j\gamma} \left[c(N-k) - e^{(k-1)\gamma}\right]}{d(k - 2)} \ , \ j = 0, 1, \ldots, k-2. \tag{5-70}$$

The next input to reach the saturation limit, as $c(N-k)$ increases, is

$u^o(N-1)$, at the point

$$c(N-k) = e^{(k-1)\gamma} + e^{(k-2)\gamma}\ d(k-2) \quad , \tag{5-71}$$

and at this point,

$$u^e(N-k+1) = e^{-(k-2)\gamma} \quad . \tag{5-72}$$

When

$$c(N-k) \geqslant e^{(k-1)\gamma} + e^{(k-2)\gamma}\ d(k-2) \quad . \tag{5-73}$$

Equation (3-10) gives

$$u^e(N) = u^e(N-1) = 1 \quad . \tag{5-74}$$

This process is continued until

$$c(N-k) = \sum_{i=0}^{k-1} e^{i\gamma} \quad . \tag{5-75}$$

In general let the values of $c(N-k)$ at which the input members $u^e(N)$, $u^e(N-1)$, ..., $u^e(N-k+2)$, $u^e(N-k+1)$ first attain the saturation limit be denoted $c_k(j)$, $j = 1, 2, ..., k$. Then $c_k(j)$ is given by,

$$c_k(j) = e^{(k-1)\gamma} + e^{(k-2)\gamma} + \ldots + e^{(k-j+1)\gamma} + e^{-(k-j)\gamma}\ d(k-j), \tag{5-76}$$

where $d(0) = 1$. When

$$c(N-k) = c_k(j), \quad \text{then } u^e(N-k+1) = e^{-(k-j)\gamma} \quad . \tag{5-77}$$

If $u^e(N-k+1)$ is plotted as a function of $c(N-k)$, a piecewise linear curve results. Figure 75 gives an example of this plot for $N = 3$ and $k = 3, 2, 1$, for the case $e^{\gamma} = 2$.

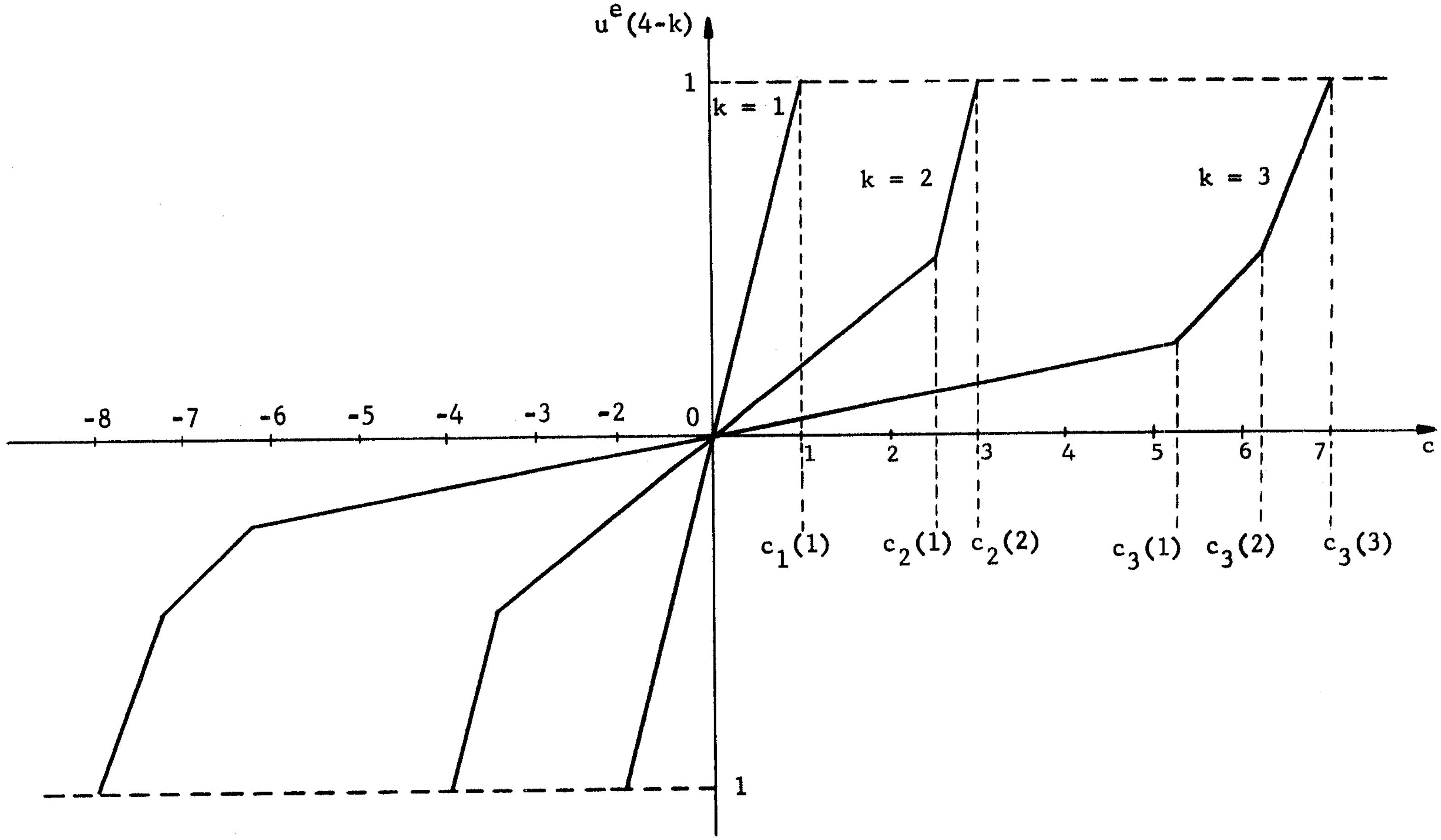


Figure 75. A plot of $u^e(1)$, $k = 3$; $u^e(2)$, $k = 2$; and $u^e(3)$, $k = 1$, for a first order plant with $N = 3$, and $e^{\delta} = 2$.

The plots of $u^e(N-k+1)$, as a function of $c(N-k)$, $k = N, N-1, \ldots, 1$, are used as the time-varying gains in the closed controller. The controller operates upon the state $c(0)$ in Γ_N with the piecewise linear gain for $k = N$ and feeds $u^e(1)$ directly into the zero order hold. One sampling period later, the gain for $k = N-1$ acts upon the state $c(1)$, giving $u^e(2)$, and so on until the regulation is complete.

The case $\lambda < 0$. This corresponds to an unstable first order plant. Since $u^o(1)$, as shown in Figure 70, page 191, is the first input to saturate, the energy optimum solution is much simpler to calculate than that for the previous case, $\lambda > 0$. Suppose k of the original N sampling periods remain to complete the regulation. Equation (3-3) gives

$$u^o(N-k+1) = c(N-k)/d(k-1) \quad , \tag{5-78}$$

so that if

$$c(N-k) \geqslant d(k-1) = \frac{1 - e^{2k\gamma}}{1 - e^{2\gamma}} \quad , \tag{5-79}$$

$$u^e(N-k+1) = 1 \quad . \tag{5-80}$$

As k decreases from $k = N$ to $k = 1$, the slope of $u^o(N-k+1)$, as a function of $c(N-k)$, decreases to unity slope at $k = 1$. The initial inputs, $u^e(1)$, $u^e(2)$, ... are therefore larger than the later inputs, as opposed to the case $\lambda > 0$. For $\lambda < 0$, the state is trying to move away from the origin, so the controller applies its main effort at once, whereas for $\lambda > 0$, the state is drifting toward the origin of its own accord, so the controller is rarely using its full effort. If c(0) is such that, for $\lambda < 0$,

$$c(0) \qquad \frac{1}{1 + e^{\gamma}} \quad , \tag{5-81}$$

Equation (4-15) shows that the plant is out of control, and no amplitude constrained input sequence can bring the state back to the origin.

<u>The case $\lambda = 0$</u>. This case has the solution, from Equation (5-63),

$$u^{o}(N-k + 1) = u^{e}(N-k + 1) = c(N-k)/k \quad . \tag{5-82}$$

The implementation in closed loop form is then exactly the same as the configurations shown in Figure 70, page 191, for the case $\lambda = 0$. Figure 70 shows how to implement the controller for the minimum fuel input sequence, but when $\lambda = 0$, the input sequences for minimum energy and minimum fuel are identical. Note that the time-varying gain in Figure 70, and Equation (5-42), is apparently different from that of Equation (5-82), but, in this latter case, k takes on the values N, N-1, ..., 1 sequentially, whereas in Equation (5-42), k increases from 0 to N-1.

<u>Second Order Systems</u>

The closed loop control for second order systems is considered only for two cases: plants with tuned complex poles, and plants with integration.

<u>Plants with tuned complex poles</u>. Suppose N-k sampling periods have elapsed since the regulation process started, so that only k

sampling periods remain to complete the task of bringing the state to the origin. Equation (5-10) shows that the next input, $u(N-k+1)$, depends only on the first component, $c_1(N-k)$, of the state $c(N-k)$. Thus

$$c_1(N-k) = u(N-k+1) + \sum_{\substack{j=3 \\ j \text{ odd}}}^{k} u(N-k+j)(-1)^{(j-1)/2} e^{(j-1)aT} . \tag{5-78}$$

The closed loop control is obtained by comparison with the known closed loop control of a first order system. Consider Equation (4-48), which gives the deadbeat constraint for a plant of the form

$$G_p(s) = \frac{1}{s+2a} \tag{5-79}$$

as,

$$c = u(1) + \sum_{j=2}^{k} u(j) e^{(j-1)2aT} . \tag{5-80}$$

The index j in Equation (5-80) runs from $j = 2$ up to $j = k$. In order to make a direct comparison of the two deadbeat constraints of Equations (5-78) and (5-80), let the upper limit in Equation (5-80) be m, where

$$m = \begin{cases} \dfrac{k-2}{2} & \text{if } k \text{ is even} \\[2ex] \dfrac{k-1}{2} & \text{if } k \text{ is odd} \end{cases} . \tag{5-81}$$

The first order plant deadbeat constraint of Equation (5-80) becomes

$$c = u(1) + \sum_{j=2}^{m} u(j) e^{(j-1)2aT} . \tag{5-82}$$

Equation (5-82) corresponds to the deadbeat constraint of the plant of Equation (5-79) when m sampling periods are allowed for the regulation. The points $c_m(j)$, j = 1, 2, ..., m, for the closed loop graphical solution of $u^e(1)$, are, from Equation (5-76),

$$c_m(j) = e^{(m-1)\gamma} + \ldots + e^{(m-j+1)\gamma} + e^{-(m-j)\gamma} \; d(m-j) \; , \qquad (5\text{-}83)$$

where $\gamma = 2aT$. When

$$c = c_m(j), \quad \text{then } u^e(1) = e^{-(m-j)\gamma} \; . \qquad (5\text{-}84)$$

The differences between Equations (5-78) and (5-82), the different notation and the alternating signs of the input members, do not prevent Equations (5-83) and (5-84) from giving $u^e(N-k+1)$ as a function of $c_i(N-k)$. For example, suppose $e^{\gamma} = e^{2aT} = 2$. Figure 75, page 203, shows the input sequence $u^e(N-k+1)$ for the first order system of Equation (5-79), when N = 3 and k = 3, 2, 1. These same plots may be used for the second order plant with tuned complex poles, given by Equations (5-8) and (5-9) as,

$$G_p(s) = \frac{1}{\left[s + a + j\frac{\pi}{2T}\right]\left[s + a - j\frac{\pi}{2T}\right]} \; , \qquad (5\text{-}85)$$

when N is either five or six. The closed loop controller, therefore, uses the function corresponding to k = 3 for two sampling periods if N = 6, or one sampling period if N = 5. This piecewise linear gain is then changed to the function corresponding to k = 2 for the next two sampling periods. Finally, the unit gain gives the last two inputs, $u^e(5)$ and $u^e(6)$ if N = 6, or $u^e(4)$ and $u^e(5)$ if N = 5.

The cases $a = 0$ and $a < 0$ may be solved in exactly the same manner. The time-varying gains have the same form as the corresponding first order system of Equation (5-79).

<u>Plants with integration</u>. The closed loop control of second order plants of the form,

$$G_p(s) = \frac{1}{s(s+\lambda)} \tag{5-86}$$

$$G_p(s) = \frac{1}{s^2} \tag{5-87}$$

can be derived by considering the sets M_k and Γ_k, $k = N, N-1, \ldots, 2$. It will be shown that the implementation of the true optimum closed loop controller would not be a practical proposition. However, the consideration of the optimum controller leads directly to a practical suboptimum controller.

A. <u>True optimum closed loop control</u>. The requirements for the optimum closed loop controller will be considered for the plant $1/s^2$. The controller requirements for the plant of Equation (5-86) are quite similar. Figure 76 shows the sets Γ_3 and M_3 for the plant

$$G_p(s) = \frac{1}{s^2} \quad . \tag{5-88}$$

The set M_3 is shown as the dashed parallelogram. Because of symmetry, only intital states with $c_2 \leqslant 0$ need be considered. Suppose, with $N = 3$, the initial state $\underline{c}(0)$ lies in M_3. Then $u^e(1) = u^o(1)$. If $\underline{c}(0)$ lies such that $u^o(1) > 1$, the conditions of Theorem 3 always being

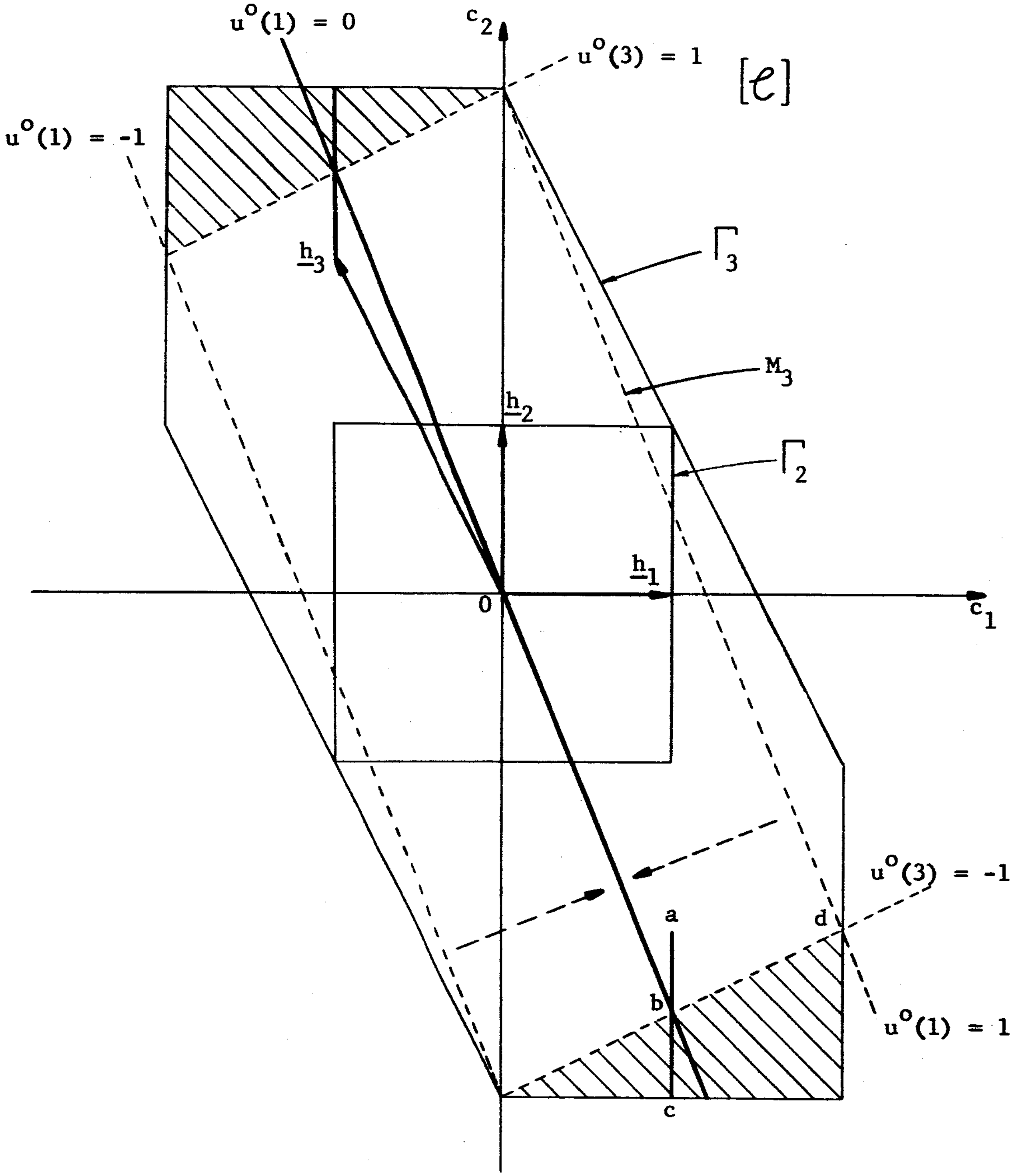


Figure 76. The sets Γ_3 and M_3 for the plant $1/s^2$ showing how to obtain $u^e(1)$ in a closed loop manner.

satisfied,

$$u^e(1) = \text{sgn. } u^o(1) = 1 \quad . \tag{5-89}$$

In the cross-hatched region, $u^o(3) < -1$, so that

$$u^e(3) = \text{sgn. } u^o(3) = -1 \quad . \tag{5-90}$$

Setting $u^e(3) = 1$ gives the new deadbeat constraint as

$$\underline{c}(0) - \underline{h}_3 = u(1)\,\underline{h}_1 + u(2)\,\underline{h}_2 \quad . \tag{5-91}$$

Equation (5-91) gives $u^e(1)$ and $u^e(2)$ uniquely.

The closed loop procedure to obtain $u^e(1)$ for N = 3 is therefore as follows. If the state lies in the cross-hatched region, $u^e(1)$ is the horizontal distance of the state from the line ac shown in Figure 76. If the state does not lie in the cross-hatched region, it follows that

$$u^e(1) = \text{sat. } u^o(1) \quad , \tag{5-92}$$

where $u^o(1)$ is obtained in the usual closed loop form, as a linear vector gain operating on $\underline{c}(0)$. The vector gain is the first row of the matrix $\left[I + HH^t\right]^{-1}$, where,

$$H = \left[\underline{h}_3\right] \quad . \tag{5-93}$$

Graphically, $u^o(1)$ is the length of the projection of the state $\underline{c}(0)$ onto the line ob, divided by the distance bd. The dashed arrows in Figure 76 show the directions of projection, perpendicular to the line ob. The line ob is simply the line $u^o(1) = 0$.

Even for N = 3, the implementation of the closed loop controller would be a difficult task, since the choice of feedback gain would depend on whether the state was in the cross-hatched region or not. It is found, by a procedure similar to that for the case N = 3, that when N = 4, the number of regions where the projection is different increases from two to three. In each region a different feedback would be necessary. The problem of implementation is not particularly that of implementing the different feedback gains, but rather the difficulty of deciding which gain is to be used. Figure 77 shows the sets M_3, Γ_4 and M_4 for the plant $1/s^2$. The different slopes of the cross-hatched differentiate the regions where the feedback strategy is different. The lines onto which the state is projected are shown as the dashed lines in Figure 77.

In general, for the plant of Equation (5-88), if the settling time is to be N sampling periods, the state must be identified as lying in one of N regions before the appropriate feedback can be selected. This would be quite impractical to implement.

B. Suboptimum closed loop control. Figures 76 and 77 show that the regions in the sets Γ_3 and Γ_4, bounded respectively by the lines $u^o(3) = \pm 1$ and $u^o(4) = \pm 1$, constitute the major portion of these sets. This is also true of the plant of Equation (5-86). It is therefore suggested that the feedback be given by Equation (5-92) for all initial states.

This feedback is implemented by the controller configuration of Figure 78. The time-varying gains, $\underline{f}(j)$, j = 1, 2, ..., N, are the same as those used in the linear minimum energy controller:

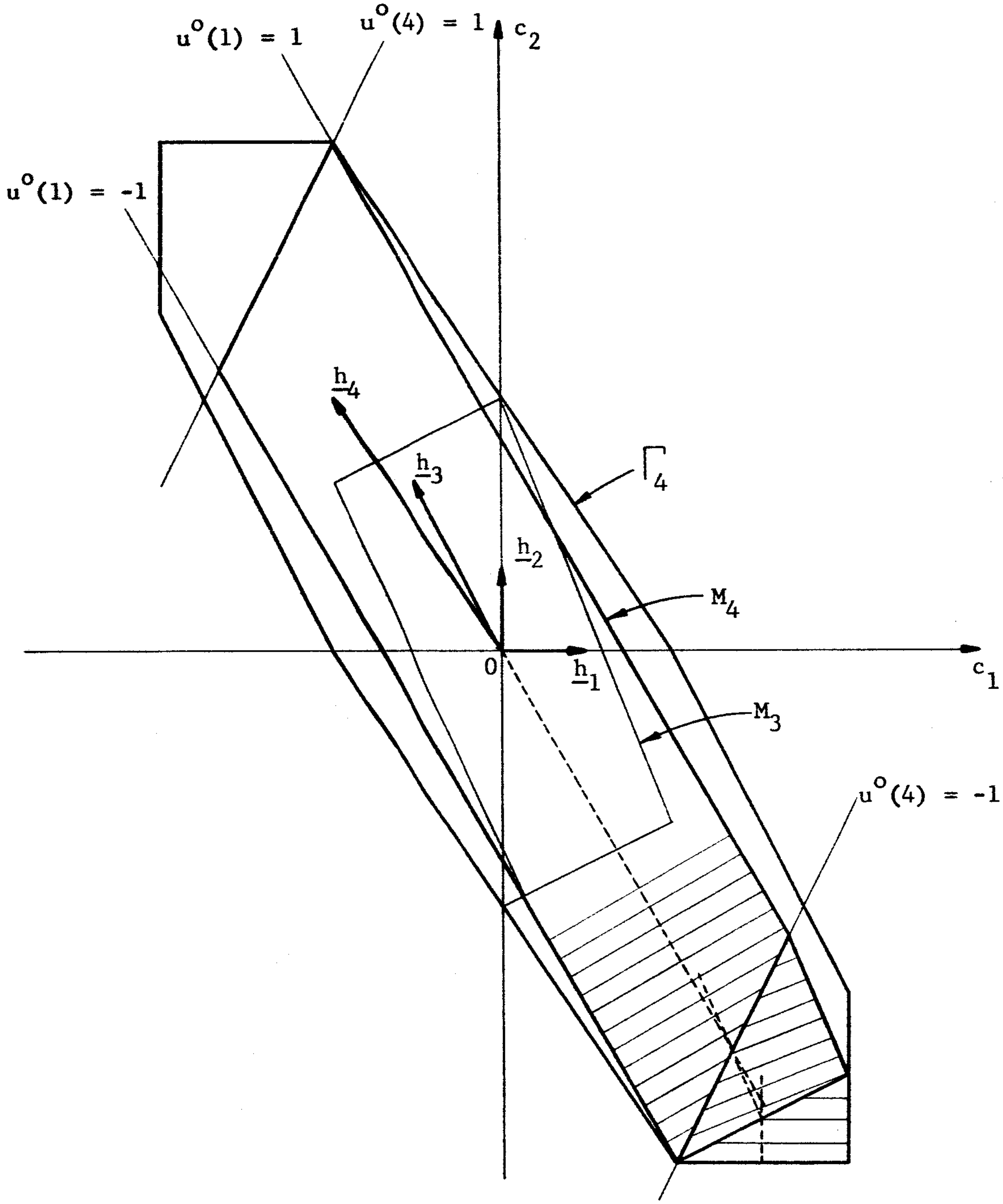


Figure 77. The regions in Γ_4 where the closed loop controller requires different stategies for the plant $1/s^2$.

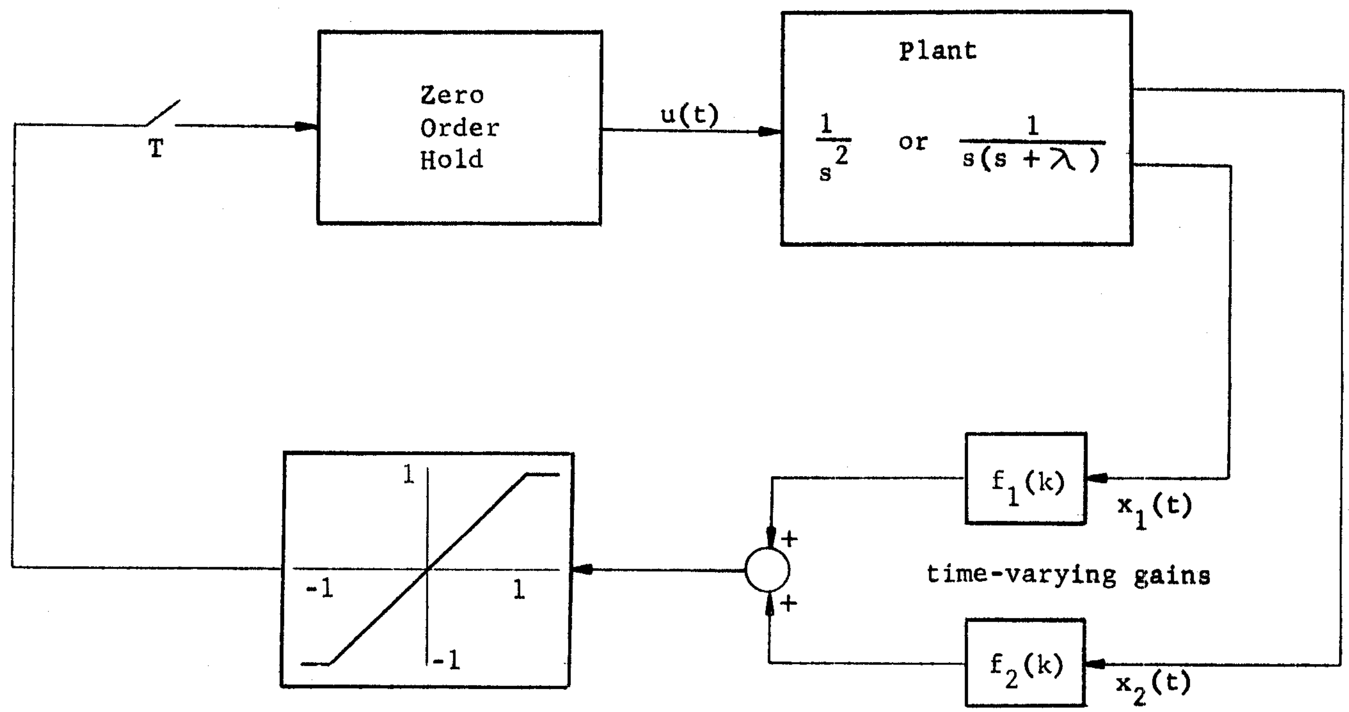


Figure 78. The system configuration for the suboptimum minimum energy control of the plants $1/s^2$ or $1/s(s + \lambda)$.

$$\underline{f}(j+1) = \begin{bmatrix} f_1(j+1) \\ \\ f_2(j+1) \end{bmatrix} = \underline{r}_1^t \left[CC^t\right]^{-1}, \; j = 0, 1, \ldots, N\text{-}n\text{-}1 , \tag{5-94}$$

where

$$C = \left[\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_{N-j}\right] , \tag{5-95}$$

and

$$\underline{f}(j) = \underline{e}\left[\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_n\right]^{-1} , \; j = N\text{-}n, \ldots, N , \tag{5-96}$$

where $\underline{e}$ is the 1 x n row vector given by

$$\underline{e} = \left[1, 0, \ldots, 0\right] . \tag{5-97}$$

VIII. CLOSED LOOP CONTROL FOR MINIMUM FUEL WITH INPUT SATURATION

First Order Systems

The open loop control of the first order minimum fuel system with input saturation is discussed in Chapter IV. Figure 53, page 138, illustrates a graphical method of obtaining the optimum input sequence. The first order plant being given by,

$$G_p(s) = \frac{1}{s+\lambda} , \tag{5-98}$$

consider the three cases, $\lambda > 0$, $\lambda < 0$, and $\lambda = 0$.

The case $\lambda > 0$. Suppose (N-k)T seconds have elapsed since the controller generated $u^f(1)$ from c(0) at time t = 0. In order that the controller be able to take the state c(N-k) into the origin in the remaining kT seconds, c(N-k) must lie in Γ_k; i.e., from Equation (4-8),

$$\left|c(N-k)\right| \leqslant \sum_{j=0}^{k} e^{(j-1)\lambda T} \quad . \tag{5-99}$$

Assuming Equation (5-99) is satisfied, the closed loop control is given, compare Figure 53, page 138, as follows. If

$$\left|c(N-k)\right| \leqslant \sum_{j=1}^{k} e^{(j-1)\lambda T} \quad , \tag{5-100}$$

note the lower index, j = 1, on the summation, then

$$u^f(N-k+1) = 0 \quad . \tag{5-101}$$

If, assuming without loss of generality than $c(N-k) > 0$,

$$\sum_{j=1}^{k} e^{(j-1)\lambda T} \leqslant c(N-k) \leqslant \sum_{j=0}^{k} e^{(j-1)\lambda T} \tag{5-102}$$

then,

$$u^f(N-k+1) = \text{sat.}\left[c(N-k) - \sum_{j=1}^{k} e^{(j-1)\lambda T}\right] \quad . \tag{5-103}$$

<u>The case $\lambda < 0$.</u> In this case, the longest invariant vector is $\underline{h}_1$, so that, for all k, k = N, N-1, ..., 1, the closed loop control is given as,

$$u^f(N-k+1) = \text{sat. } c(N-k) \quad . \tag{5-104}$$

<u>The case $\lambda = 0$.</u> Since the invariant vectors in this case are all of unit length, there are an infinite number of optimum input sequences, and therefore controllers, which can give $\underline{u}^f$. The simplest closed loop controller is the one which obeys Equation (5-104).

The implementation of the controllers. The implementation of the controllers is very straightforward if $\lambda \leqslant 0$. Equation (5-104) implies that the state $c(N-k)$, $k = N, N-1, \ldots, 1$, is fed directly into a fixed saturation nonlinearity, with unit gain over its linear region. The output of this nonlinearity is then fed directly into the sample-hold device.

When $\lambda > 0$ the controller can be imagined as a variable dead zone, whose input is the plant state. The dead zone would be symmetrical, and the amount of dead zone would depend on the amount of time remaining for regulation. For example, when $0 \leqslant t < T$, the dead zone, $z(0)$, is given by

$$- \sum_{j=1}^{N} e^{(j-1)\lambda T} \leqslant z(0) \leqslant \sum_{j=1}^{N} e^{(j-1)\lambda T} . \qquad (5\text{-}105)$$

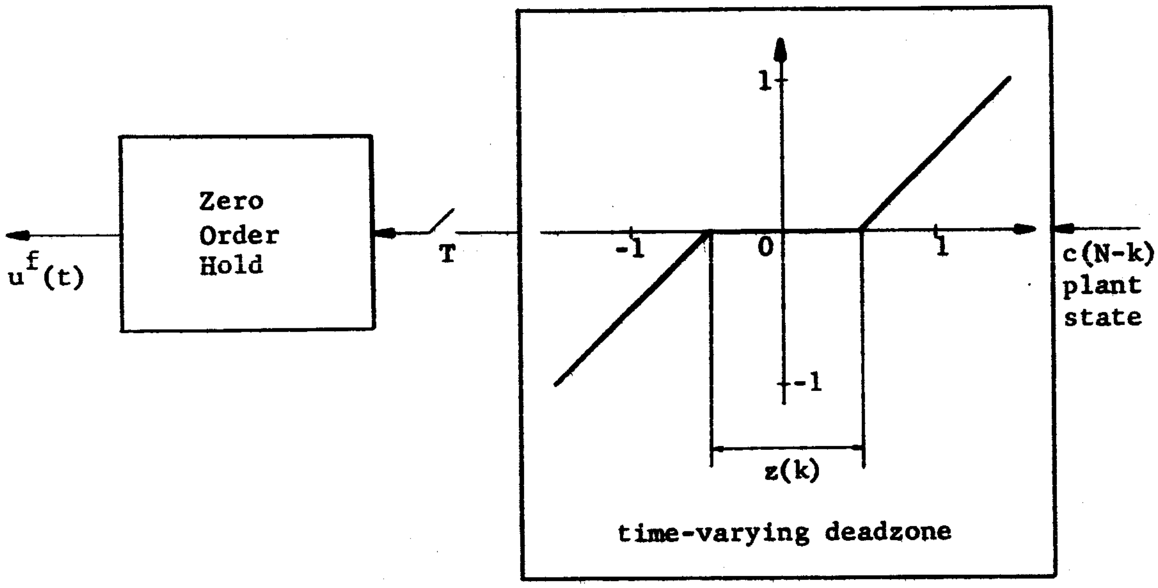
In general, for $(N-k+1)T > t \geqslant (N-k)T$, the dead zone, $z(N-k)$, is given by

$$- \sum_{j=1}^{k} e^{(j-1)\lambda T} \leqslant z(N-k) \leqslant \sum_{j=1}^{k} e^{(j-1)\lambda T} \qquad (5\text{-}106)$$

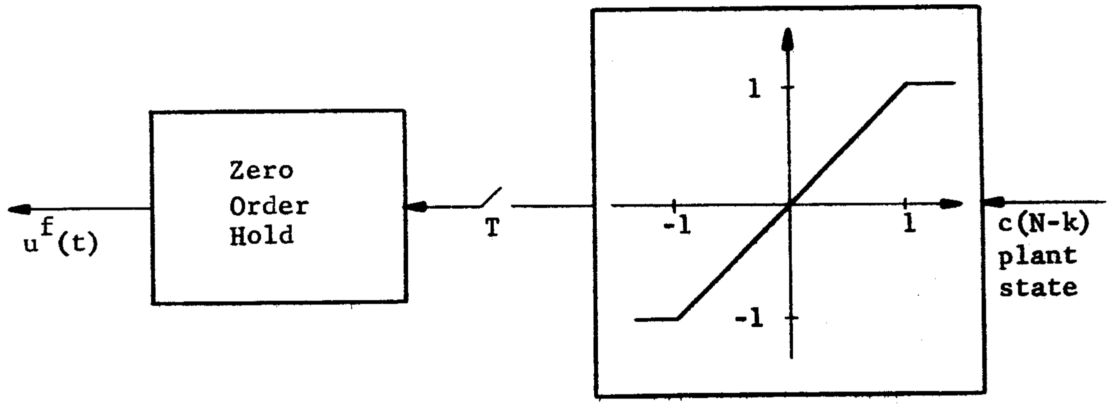
Figure 79 shows the controller configuration for the cases $\lambda > 0$ and $\lambda \leqslant 0$.

Second Order Systems

Second order systems with tuned complex poles. This case has already been discussed in Chapter IV. The controller configurations



a. The case $\lambda > 0$



b. The case $\lambda \leqslant 0$

Figure 79. The optimum closed loop controller configurations for first order systems with minimum fuel consumption.

are exactly the same as those shown in Figure 79. In the case of a plant corresponding to $\lambda > 0$, each dead zone is retained for two sampling periods, except perhaps at the start of the regulation, when, if N is odd, the first dead zone, z(0), is changed after only one sampling period. Figure 66, page 169, gives an example of the dead zone non-linearity when the settling time is given as either five or six sampling periods.

Second order systems with integration. For plants of the form of Equations (5-86) and (5-87), the closed loop control is obtained by generalizing Equation (4-37). If the state $\underline{c}(N-k)$ is in the set Q_k, $k = N, N-1, \ldots, 2$,

$$u^f(N-k+1) = 0 \ . \tag{5-107}$$

If $\underline{c}(N-k)$ is in Γ_k, but not in Q_k, then

$$u^f(N-k+1) = \text{sat.}\ \mu \ , \tag{5-108}$$

where μ is the smallest number in absolute value such that $\underline{c}(N-k) - \mu\ \underline{h}_1$ lies on Q_k. The sets Q_k are not unique for plants with integration, however, the simplest controller configuration is obtained when Q_k is constructed as follows. In $\mathcal{C}$-space, let the set Q_k, for $c_2(N-k) > 0$, be the line joining the set of points:

$$\underline{h}_k,\ \underline{h}_k + \underline{h}_{k-1},\ \underline{h}_k + \underline{h}_{k-2},\ \ldots,\ \sum_{j=0}^{k-2} \underline{h}_{k-j} \ . \tag{5-109}$$

For $c_2(N-k) < 0$, the set Q_k is defined by symmetry. Figure 64, page 165, for example, shows the sets Q_4 and Q_3 for the plant $1/s^2$. The closed loop controller therefore requires the use of time-varying piecewise linear gains. Figure 80 shows the actual configuration of the optimum controller. Over the last two sampling periods the gain remains fixed corresponding to the set Q_2, so that the saturation nonlinearity only receives the c_1 component of the plant state.

General second order systems. If the sequence of sets, Q_N, Q_{N-1}, ..., Q_2 can be found, see Chapter IV, page 161, the optimum closed loop controller in general would need to use two time-varying piecewise linear gains in the configuration of Figure 73, page 197. The same logic would also be necessary and would be followed by a saturation nonlinearity.

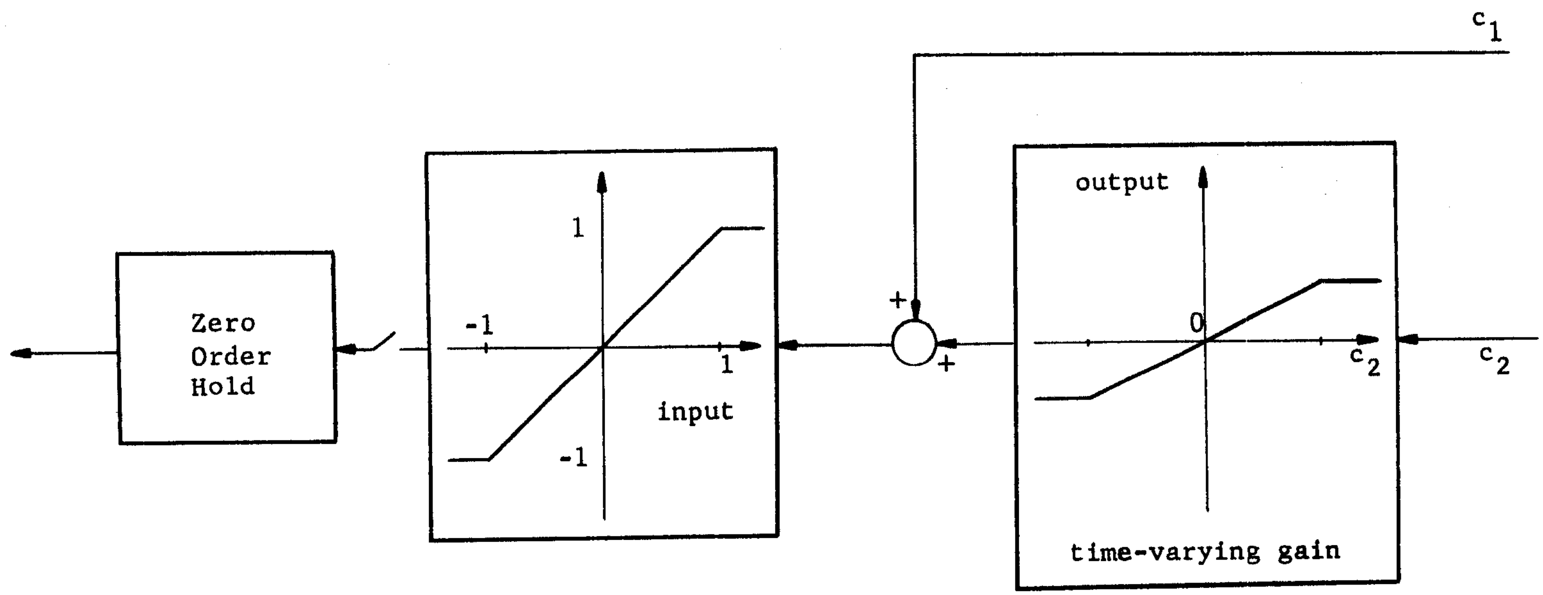


Figure 80. The closed loop optimum controller for the minimum fuel problem with input saturation for second order plants with integration.

CHAPTER VI

SUMMARY AND CONCLUSIONS

One common approach to the problem of designing control systems uses the analog computer to simulate the plant. By a process of intuition and experiment a controller may be designed which meets certain general specifications on the performance of the overall system. While this approach is frequently quite successful, it basically tailors the controller to suit existing hardware and techniques of analysis. Consequently, when the controller has been constructed and evaluated, there is often no clear indication as to how it might be further improved. While the results of theoretical analyses, with their accompanying simplifications and somewhat arbitrary performance criteria, may not be directly applicable to real systems, in some cases sufficiently realistic cost functions can be mathematically formulated and the theoretically optimum input sequence defined. Any optimum controller which results from the theoretical analysis may then be judged by balancing such factors as cost and reliability against the economic advantages of attaining an optimum system.

The discrete regulator is assuming an important role with the increasing tendency of modern systems to use digital techniques. The discrete deadbeat regulator, designed to minimize the energy and fuel cost functions, has taken on a new practical significance with the sophisticated requirements for the guidance and control of space vehicles. Although the problem of finding the optimum bounded input sequence which minimizes these two cost functions can be formally solved by using non-linear and linear programming techniques, such methods are intrinsically unable to suggest either improvements to existing control systems, or novel and simpler hardware to implement the optimum sequence. It has therefore been the aim of this research to make an investigation of the characteristics of the optimum input sequence, so that the controller could be designed around the input sequence, rather than the sequence around the controller.

I. SUMMARY OF THE APPROACHES USED AND THE RESULTS OBTAINED

When the plant is driven by the output of a zero order hold, its equation of motion can be conveniently described by a first order vector difference equation. In order to avoid having to choose a particular state space in which to represent the plant and its state, an alternate space, the canonical vector space ($\mathcal{C}$ -space) has been defined. Formulating the discrete deadbeat regulation in $\mathcal{C}$-space, via the invariant vectors, has the advantage that the properties of the input sequence need only be considered with reference to the poles of the plant transfer function.

The linear minimum energy and minimum fuel problems were discussed in Chapter II, and the corresponding cases with input saturation in Chapters III and IV respectively. Chapter V was mainly concerned with the implementation of the optimum control in closed loop form.

Summary of the Minimum Energy Problem

The linear minimum energy input sequence was found, for the general n-th order plant, by using only elementary differential calculus. The generalized energy cost function, which, by a suitable choice of the N x N matrix S, gives deadbeat control and allows the system response to be adjusted to meet various time domain specifications, was obtained by a simple extension of the ordinary energy cost function. The minimum energy control sequence for second order systems was also found by graphical techniques, using a geometrical interpretation of the optimum sequence. The open loop generation of the control requires the inversion of an n x n matrix. The closed loop implementation uses linear time varying feedback gains in the controller configuration of Figure 68, page 183.

When the saturation constraint is included, the problem of finding the optimum control is considerably more complicated. However, if the initial state, $\underline{c}$, lies in the set M_N, the linear energy equations furnish a solution, $\underline{u}^o$, which satisfies the saturation constraint. The set M_N may be obtained graphically if the order of the plant is not greater than two. If $\underline{c}$ is not in M_N, but is in the set Γ_N, it has been shown that the solution to the minimum energy regulator with input saturation amounts to finding which members of the input sequence are equal to the saturation limit. If only one member of $\underline{u}^o$ exceeds the saturation limit, then Theorem 2, page 103, guarantees that the corresponding member of $\underline{u}^e$ is equal to the limit. If more than one member of $\underline{u}^o$ exceeds the saturation limit, Theorem 3, page 120, can be used to find which of these are to be set equal to the saturation limit. Theorem 3 has two conditions which must be satisfied before a particular member may be set equal to the limit. However, the second condition, Equations (3-106) and (3-107), is, in general, quite difficult to test. It was therefore suggested that a practical open loop method for finding the constrained optimum sequence would be to use Postulate 1a, page 114, as the basis of a step by step procedure. This procedure first requires the calculation of the linear sequence $\underline{u}^o$. If any members of $\underline{u}^o$ exceed the saturation limit, Postulate 1a, or Theorem 2 if applicable, is applied to find which of these members are to be set equal to the saturation limit. Having set these members at their appropriate limits, a new deadbeat constraint results, for which a new linear optimum sequence, containing correspondingly fewer members, is calculated. Postulate 1a is then applied again if necessary. Eventually, one of two possibilities will occur. A linear optimum solution may be obtained each of whose members satisfies the saturation constraint. In this case the problem has been solved. It is possible, however, that with no more than n invariant vectors remaining to represent the latest state, there is no constrained input sequence which satisfies the corresponding deadbeat constraint. The step by step procedure has therefore erroneously set one or more of the input members at the saturation limit. It has been shown that, in general, the technique can guarantee an optimum solution only for first order systems, or for second order systems having either tuned complex poles or integration, where, if any members of the linear sequence exceed the saturation limit, they are all to be set equal to the saturation limit.

The closed loop implementation was shown to require the use of a piecewise linear time-varying gain feeding a saturation nonlinearity, when the plant is of the first order, or of the second order with tuned complex poles. Second order systems with integration were shown to require a very complex closed loop controller. A relatively simple suboptimum controller, using only the time-varying linear gains of the linear minimum energy feedback to feed the saturation nonlinearity, was suggested and is shown in Figure 78, page 213.

Summary of the Minimum Fuel Problem

The linear minimum fuel problem is approached by considering the initial state in relation to the set $S_N(f)$. For the general n-th order plant, this set is used to divide the state space into a finite number of cones. Once the state has been identified as belonging to a particular cone, the optimum sequence is easily obtained (22). The considerations involved in finding a suitable cone which contains the initial state are, in general, very involved, and have precluded investigation of any system higher than second order. First and second order systems were considered in detail. Theorem 1, page 44, gives the necessary and sufficient conditions for the uniqueness of the fuel optimum input sequence, and was utilized to investigate what combinations of plant poles and initial state give a nonunique optimum control. Open loop solutions for first order plants or second order plants with either integration or tuned complex poles are particularly simple. Other second order plants require only that the set $S_N(f)$ be constructed. The hardware required for the closed loop control of the linear minimum fuel regulator varies in complexity. For first order systems it may be only simple direct feedback, see Figure 70, page 191, while a general second order system requires a pair of linear time-varying gains providing inputs to a small logic unit. Figure 73, page 197, shows the configuration of this controller.

Open loop solutions to the minimum fuel problem with input saturation were obtained for first order systems and second order systems with tuned complex poles. It was shown that, in general, the closed loop approach is more appropriate for dealing with the saturation problem. The method suggested involves the generation of the sets Q_k, $k = N, N-1, \ldots, 3$, and these sets were obtained for second order systems with integration. More general second order systems were not investigated beyond the case $N = 4$. The closed loop controllers developed for first order systems are shown in Figure 79, page 217. Second order systems with integration were shown to require the use of a piecewise linear time-varying gain, followed by a saturation nonlinearity. Figure 80, page 220, shows the configuration. If the appropriate sets Q_k can be found, more general second order systems would incorporate two such gains followed by the logic unit and the saturation nonlinearity.

II. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

It has been shown that for first order systems and a somewhat restricted class of second order systems, practically feasible optimum closed loop controllers can be obtained, which could not have been obtained by using classical techniques. It was shown that if the poles of a second order underdamped plant can be tuned, by adjusting either the poles or the sampling period, the construction of energy and fuel optimum controllers is considerably simplified. The time-varying gain so frequently necessary would, in some cases, be a fairly costly item to produce, especially if N is large. It would therefore be of considerable value to be able to find one suboptimum time-invariant gain which could be used as a substitute. Since the closed loop feedback for deadbeat control is always constant over the last n sampling periods, the resulting linear region of control around the origin will prevent the possibility of limit-cycling, even if the plant is subject to large disturbances during the regulation process. Closed loop control for unstable plants is much simpler than that for stable plants, since the invariant vector $\underline{h}_1$ is always the longest.

On a theoretical note, the approaches used are of some interest in themselves.

The partitioning of the input sequence allowed the linear minimum energy solution to be derived in a simple manner, and, by enabling the case n = 2 and N = 4 to be studied in detail, provided a very useful method of studying the various facets of the saturation problem.

The extension to third and higher order systems presents formidable problems. Open loop procedures seem to offer more promise for the minimum energy problem, while closed loop methods, with the obvious exception of linear programming, seen more appropriate for the minimum fuel problem. The simplifications obtained when the plant has integration may continue when the higher order system has integration. Similarly, the possibility of tuning two or more pairs of complex poles exists.

Time-varying plants offer no additional theoretical obstacles: The minimum fuel and energy problems are mathematically unchanged, so that the same techniques are applicable.

REFERENCES

LIST OF REFERENCES

1. Kalman, R. E. "On the General Theory of Control Systems," Proceedings of the First International Federation of Automatic Control Congress. London: Butterworth's Publications, 1961.

2. Kalman, R. E., Y. C. Ho, and K. S. Narendra. "Controllability of Linear Dynamic Systems," Contributions to Differential Equations. New York: John Wiley and Sons, Inc., 1955. Pp. 189-213.

3. Kreindler, E., and P. E. Sarachik. "On the Concepts of Controllability and Observability of Linear Systems," Institute of Electrical and Electronic Engineers Professional Group on Automatic Control, 9(2):129-136, April, 1964.

4. Zadeh, L. A., and C. A. Desoer. Linear System Theory. New York: McGraw-Hill Book Company, Inc., 1963.

5. Tou, J. T. Modern Control Theory. New York: McGraw-Hill Book Company, Inc., 1964.

6. Dorf, R. C. Time Domain Analysis and Design of Control Systems. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1965.

7. Kalman, R. E., and J. E. Bertram. "General Synthesis Procedure for Computer Control of Single Loop and Multi-Loop Linear Systems," American Institute of Electrical Engineers Transactions on Applications and Industry, 77(2):602-609, January, 1959.

8. Ragazzini, J. R., and G. F. Franklin. Sampled-Data Control Systems. New York: McGraw-Hill Book Company, Inc., 1958.

9. Tou, J. T. Digital and Sampled-Data Control Systems. New York: McGraw-Hill Book Company, Inc., 1959.

10. Kuo, B. C. Analysis and Synthesis of Sampled-Data Control Systems. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1963.

11. Hung, J. C. "Sampled-Data Control Systems with Two Digital Processing Units," American Institute of Electrical Engineers Transactions on Applications and Industry, 79(2):292-298, September, 1960.

12. Revington, A. M., and J. C. Hung. "A General Theory on Multivariable Sampled-Data Control Systems," Proceedings of the National Winter Convention on Military Electronics in Los Angeles February 5 to February 7, 3:20-28, February, 1964.

13. Cadzow, J. A. "A Study of Minimum Norm Control for Sampled-Data Systems," Joint Automatic Control Conference, 8:545-550, 1965.

14. Desoer, C. A., and J. Wing. "An Optimal Strategy for a Saturating Sampled-Data System," Institute of Radio Engineers Professional Group on Automatic Control, 6(2):5-15, February, 1961.

15. Revington, A. M., and J. C. Hung. "Design of Minimum Energy Discrete-Data Control Systems," NASA-CR-60604. Washington: Office of Technical Services, Department of Commerce, 1965.

16. Bertram, J. E., and P. E. Sarachik. "On Optimal Computer Control," Proceedings of the First International Federation of Automatic Control Congress. London: Butterworth's Publications, 1961.

17. Kalman, R. E., and R. W. Koepcke. "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indices," American Society of Mechanical Engineers, 80D:1800-1826, November, 1958.

18. Revington, A. M., and J. C. Hung. "Design of Time-Weighted Minimum Energy Discrete-Data Control Systems," NASA-CR-60512. Washington: Office of Technical Services Department of Commerce, 1965.

19. Penrose, R. "A Generalized Inverse for Matrices," Proceedings of the Cambridge Philosophical Society, 51:406-413, 1955.

20. Penrose, R. "On Best Approximate Solutions of Linear Matrix Equations," Proceedings of the Cambridge Philosophical Society, 52:17-19, 1956.

21. Yuji, I. "On the Generalized Inverse of an Incidence Matrix," Journal of the Society of Industrial and Applied Mathematics, 13(13):827-836, September, 1965.

22. Lee, C. T., and C. A. Desoer. "Fuel Optimal Control for Pulse-Amplitude-Modulated Time-Invariant Sampled-Data System," Journal of Electronics and Control, 15(4):363-381, 1963.

23. Kalman, R. E. "Optimum Non-Linear Control of Saturating Systems by Intermittent Action," Institute of Radio Engineers WESCON Record, 1:130-135, 1957.

24. Desoer, C. A., and J. Wing. "A Minimum Time Discrete System," Institute of Radio Engineers Professional Group on Automatic Control, 6(5):111-124, May, 1961.

25. Desoer, C. A., and J. Wing. "Minimal-Time Regulator Problem for Sampled-Data Systems (General Theory)," Journal of the Franklin Institute, 271:208-228, September, 1961.

26. Kurzweil, F., Jr. "The Analysis and Synthesis of Non-Linear Continuous and Sampled-Data Systems Involving Saturation," Stanford Electronic Laboratories Technical Report No. 2101-1. November, 1959.

27. Nagata, A., S. Kadama, and S. Kumagai. "Time Optimal Discrete Control with Bounded State Variable," Institute of Electrical and Electronic Engineers Professional Group on Automatic Control, 10(2):155-164, April, 1965.

28. Tou, J. T. "Synthesis of Discrete Systems Subject to Control Signal Saturation," Journal of the Franklin Institute, 277: 401-413, May, 1964.

29. Torng, H. C. "Optimization of Discrete Systems Through Linear Programming," Journal of the Franklin Institute, 278:28-44, July, 1964.

30. Koepcke, R. W. "A Solution to the Sampled, Minimum Time Problem," Joint Automatic Control Conference, 6:94-99, 1963.

31. Ho, Y. C. "Solution Space Approach to Optimal Control Problems," American Society of Mechanical Engineers, 83D:53-58, March, 1961.

32. Kim, M. "Optimum Discrete Control with Amplitude and Energy Constrained Control Function," Joint Automatic Control Conference, 8:539-544, 1965.

33. Fegley, K. A. "Designing Sampled-Data Control Systems by Linear Programming," Institute of Electrical and Electronic Engineers Transactions on Applications and Industry, 83(72):198-200, May, 1964.

34. Ho, Y. C. "A Successive Approximation Technique for Optimal Control Systems Subject to Input Saturation," American Society of Mechanical Engineers, 84D:33-40, March, 1962.

35. Ho, Y. C., and P. B. Brentani. "On Computing Optimal Control with Inequality Constraints," Society of Industrial and Applied Mathematics, 1A(3):319-348, December, 1963.

36. Stubberud, A. R., and J. M. Swiger. "Minimum Energy Control of a Linear Plant with Magnitude Constraint on the Control Input Signals," Joint Automatic Control Conference, 8:398-406, 1965.

37. Kalman, R. E., and R. W. Koepcke. "The Role of Digital Computers in the Dynamic Optimization of Chemical Reactions," Proceedings of the Western Joint Computer Conference, 1959. Pp. 107-116.

38. Åström, K. J., R. W. Koepcke, and F. Tung. "On the Control of Linear Discrete Dynamic Systems with Quadratic Loss," International Business Machines Research Report RJ-222. September, 1962.

39. Deley, G. W., and G. F. Franklin. "Optimum Bounded Control of Linear Sampled-Data Systems with Quadratic Loss," American Society of Mechanical Engineers, 87D(1):135-141, March, 1965.

40. Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems," American Society of Mechanical Engineers, 82D: 35-45, March, 1960.

41. Smith, G. L. "The Scientific Inferential Relationships Between Statistical Estimation, Decision Theory, and Modern Filter Theory," Joint Automatic Control Conference, 8:350-359, 1965.

42. Greensite, A. L. "Dynamic Programming and the Kalman Filter Theory," Proceedings of the National Electronics Conference, 20:601-605, 1964.

43. Hadley, G. Non-Linear and Dynamic Programming. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1964.

44. Greville, T. N. E. "The Pseudoinverse of a Rectangular or Singular Matrix and Its Application to the Solution of Systems of Linear Equations," Society of Industrial and Applied Mathematics Review, 1(1):38-43, January, 1959.

45. Revington, A. M., and J. C. Hung. "Geometrical Interpretation and Graphical Solution to Minimum Energy Discrete-Data Control," NASA-CR-68498. Springfield, Virginia: Clearinghouse for Federal Scientific and Technical Information, 1966.

46. Hadley, G. Linear Algebra. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1961.

47. Nelson, W. L. "Optimal Control Methods for On-Off Sampling Systems," American Society of Mechanical Engineers, 83D:91-100, March, 1962.

48. Hung, J. C., and A. M. Revington. "Minimum Energy Design for Discrete-Data Control Systems," Proceedings of the International Federation of Automatic Control Tokyo Symposium, August, 1965.

APPENDICES

APPENDIX A

SAMPLED-DATA SYSTEMS AND THE INVARIANT VECTORS

I. INTRODUCTION

The Appendix begins with a discussion of the plant and its response to pulse-amplitude-modulated control signals, following which the discrete deadbeat regulator problem is formulated in terms of the canonical vectors. Whether or not there is a solution to the problem depends on the controllability of the plant, and the necessary conditions for the existence of a solution in terms of the controllability of the plant are given. When considering control sequences that are limited in amplitude, additional information is needed. The information is considered in terms of the set Γ_N.

The regulator problem can be clarified and its solution simplified if the formulation with the canonical vectors is replaced by a formulation using the invariant vectors. These vectors are introduced and tabulated for first and second order plants. Finally the regulator problem is reformulated with these invariant vectors, the problem being considered in the canonical space rather than the state space.

II. THE PLANT

The n-th order linear plant is described by the matrix differential equation,

$$\dot{\underline{x}}(t) = A\,\underline{x}(t) + \underline{d}\,u(t) \quad . \tag{A-1}$$

The plant output c(t) shown in Figure 3, page 5, is a linear combination of the state variables. The solution of this differential equation is

$$\underline{x}(t) = G(t - t_o)\,\underline{x}(t_o) + \int_{t_o}^{t} G(t - \tau)\underline{d}u(\tau)d\tau \;, \quad t \geqslant t_o. \tag{A-2}$$

In general, for the time invariant plant, t_o may be taken to be zero. The transition matrix, G(t), may be found by several methods. One convenient formula is,

$$G(t) = \mathcal{L}^{-1}\left[\left[sI - A\right]^{-1}\right] \quad , \tag{A-3}$$

where I is the n x n identity matrix and $\mathcal{L}^{-1}$ denotes the inverse Laplace transformation.

Consider the case when the plant is subjected to pulse-amplitude-modulated inputs (8, 9, 10). Suppose

$$u(t) = u(1) = \text{constant}, \quad 0 \leqslant t < T \; . \tag{A-4}$$

Then

$$\underline{x}(t) = G(t)\,\underline{x}(0) + u(1)\,\underline{h}(t) \quad , \tag{A-5}$$

where

$$\underline{h}(t) = \int_{0}^{t} G(t - \tau)\underline{d}\,d\tau = \int_{0}^{t} G(\tau')\underline{d}\,d\tau' \quad . \tag{A-6}$$

After T seconds the solution is

$$\underline{x}(T) = G(T)\,\underline{x}(0) + u(1)\,\underline{h}(T) \quad . \tag{A-7}$$

After $(k + 1)T$ seconds the solution is

$$\underline{x}(\overline{k + 1T}) = G(T)\ \underline{x}(kT) + u(k + 1)\ \underline{h}(T) \quad , \tag{A-8}$$

where $u(k + 1)$ is a constant input over $kT \leqslant t < (k + 1)T$ as shown in Figure 1, page 2. Some useful properties of $G(t)$ and $\underline{h}(t)$ are given below. Letting t_1 and t_2 be arbitrary real numbers and k be an integer,

$$G(0) = I \quad , \tag{A-9}$$

$$\underline{h}(0) = 0 \quad , \tag{A-10}$$

$$G(t_1 + t_2) = G(t_1)G(t_2) \quad , \tag{A-11}$$

$$G^{-1}(t_1) = G(-t_1) \quad , \tag{A-12}$$

$$G^{-k}(t_1) = G(-kt_1) \quad , \tag{A-13}$$

and

$$\underline{h}(t_1 + t_2) = G(t_1)\underline{h}(t_2) + \underline{h}(t_1) \quad . \tag{A-14}$$

Beginning with the initial state $\underline{x}(0)$ and using Equation (A-8) repeatedly gives

$$\underline{x}(1) = G(T)\underline{x}(0) + u(1)\underline{h}(T) \quad , \tag{A-15}$$

$$\begin{aligned} \underline{x}(2) &= G(T)\underline{x}(1) + u(2)\underline{h}(T) \quad , \\ &= G(2T)\underline{x}(0) + u(1)G(T)\underline{h}(T) + u(2)\underline{h}(T) \quad , \end{aligned} \tag{A-16}$$

where $\underline{x}(k) = \underline{x}(kT)$ for notational convenience. Finally,

$$\underline{x}(N) = G(NT)\underline{x}(0) + \sum_{j=1}^{N} u(j)G(\overline{N - jT})\underline{h}(T) \quad . \tag{A-17}$$

Equation (A-17) gives the solution to Equation (A-1) when the input is of the form of Figure 1, page 2. The solution at instants of time other than $t = kT$, $k = 1, 2, \ldots$ can be found (5, 6), but such considerations will not be needed.

III. THE DEADBEAT REGULATOR

The deadbeat regulator requires that $\underline{x}(N) = 0$. The condition for $\underline{x}(N) = 0$ can be obtained by premultiplying Equation (A-17) by $G(-NT)$. Then with the use of Equations (A-9), (A-11) and (A-13), there results

$$\underline{x}(0) = \sum_{j=1}^{N} -G(-jT)\underline{h}(T)u(j) \quad . \tag{A-18}$$

The canonical vectors $\underline{r}_j$, $j = 1, 2, \ldots,$ are defined as (7)

$$\underline{r}_j = -G(-jT)\underline{h}(T) \quad . \tag{A-19}$$

With the use of Equation (A-14),

$$\underline{r}_j = \underline{h}(-jT)-\underline{h}(-(j - 1)T) \quad . \tag{A-20}$$

From Equation (A-18) a necessary and sufficient condition (1, 23, 25) that the state of the plant can be brought to the origin in N sampling periods by the input sequence $u(1), u(2), \ldots, u(N)$, is that the initial state $\underline{x}(0)$ be given by

$$\underline{x}(0) = \sum_{j=1}^{N} u(j)\, \underline{r}_j \quad . \tag{A-21}$$

Equation (A-21) is fundamental to the deadbeat regulator problem.

Controllability

Pulse-amplitude-modulated plants described by the difference equation, Equation (A-8), are defined to be completely controllable if and only if the set of vectors $\underline{r}_1$, $\underline{r}_2$, ..., $\underline{r}_n$ are linearly independent.

Complete controllability is a necessary condition for there to be a solution to the deadbeat regulator problem. For the linear case, when there is no saturation constraint on the input sequence, complete controllability is also a sufficient condition. If any n of the set of canonical vectors $\underline{r}_j$, j = 1, 2, ..., N are linearly independent, they can be used as a basis for the state space. It then follows from Equation (A-21) that for any initial state $\underline{x}(0)$ there is an input sequence u(1), ..., u(N) which makes $\underline{x}(N) = 0$. In the continuous case the plant described by Equation (A-1) is completely controllable if and only if the vectors $\underline{d}$, $A\underline{d}$, ..., $A^{n-1}\underline{d}$ are linearly independent (1). Usually the introduction of the sample-hold device between the input and the plant leaves the discrete plant of Equation (A-8) completely controllable. If the plant has complex poles however, it is possible for the continuous plant to be completely controllable and for the discrete plant not to be completely controllable. It has been shown (1, 2) that the discrete plant remains completely controllable if and only if, for every eigenvalue λ of A,

$$\text{whenever} \quad \text{Re.}\,\lambda_i = \text{Re.}\,\lambda_j\,, \text{ then } \text{Im.}\left[\lambda_i - \lambda_j\right] \neq \frac{2\pi k}{T}\,, \tag{A-22}$$

where k is a positive integer. For example, if only one pair of complex poles occurs in the continuous system, so that $\lambda_1 = a + jb$, $\lambda_2 = a - jb$ are the only complex eigenvalues of A, the plant remains completely controllable if and only if

$$bT \neq k\pi \quad . \tag{A-23}$$

For second order systems this can be illustrated geometrically. From Equation (A-20),

$$\underline{r}_1 = \underline{h}(-T) \quad ,$$

$$\underline{r}_2 = \underline{h}(-2T) - \underline{h}(-T) \quad . \tag{A-24}$$

Figure 81 shows $\underline{h}(-t)$ plotted for a typical second order system with complex poles. The figure illustrates that if $T = \pi/b$, $\underline{r}_1$ and $\underline{r}_2$ are not linearly independent. It can be shown further that, in fact, all the canonical vectors lie in the same direction.

The Set Γ_N

Consider the set of all initial states that can be taken into the origin in one sampling period. This set is found by setting $\underline{x}(1) = 0$ in Equation (A-15), which gives

$$\underline{x}(0) = \underline{r}_1\, u(1) \quad . \tag{A-25}$$

If u(1) is unrestricted, all the states on the line $u(1)\underline{r}_1$ can be brought into the origin in one sampling period. If $|u(1)| \leqslant 1$, only states lying along the vector $\underline{r}_1$ or $-\underline{r}_1$ can be brought into the origin in one sampling period. Similarly the set of all states that can be taken to the origin in two sampling periods or less, is the set of all states that can be taken to the state $u(2)\underline{r}_1$ in one sampling period.

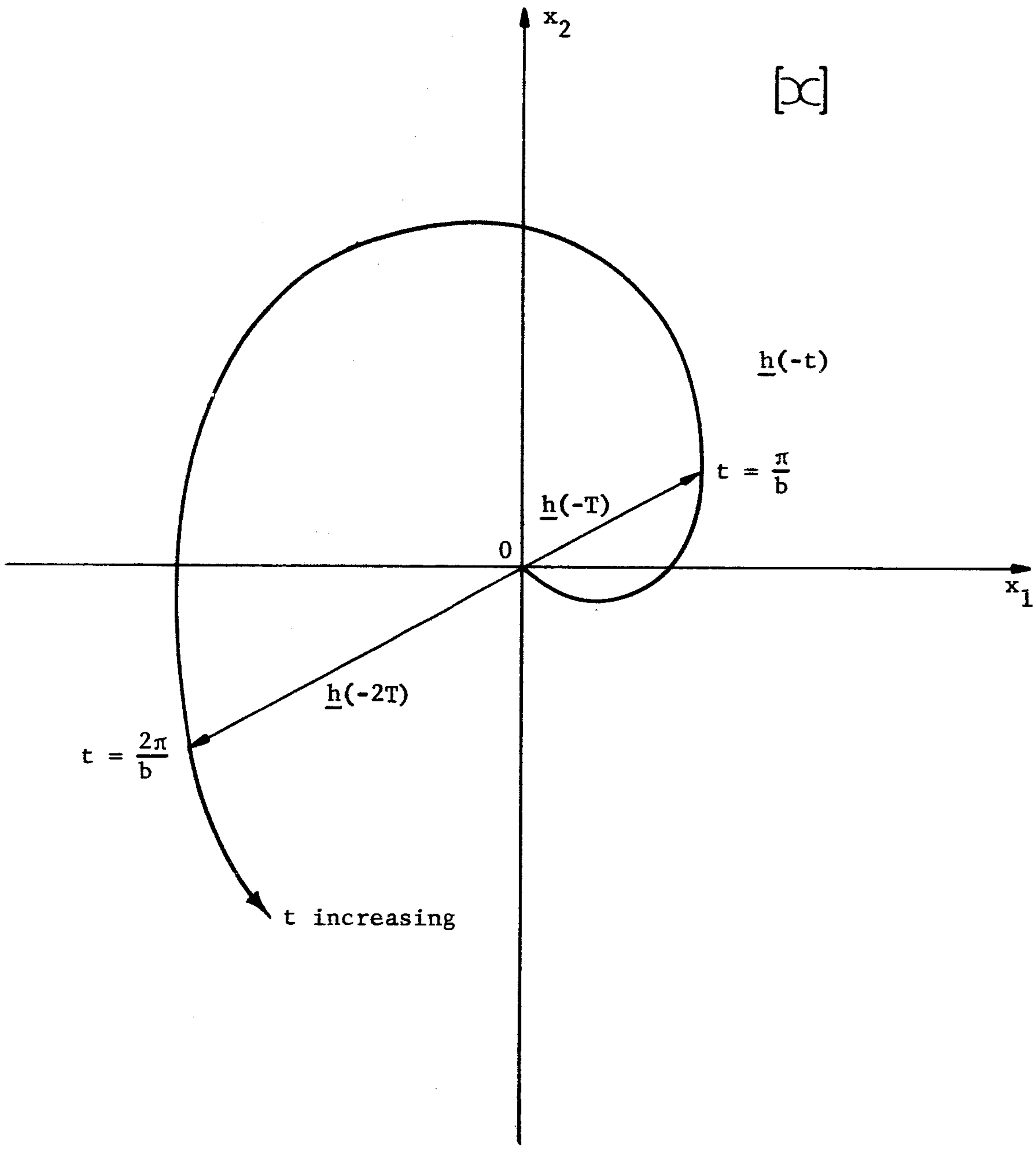


Figure 81. Loss of controllability is possible with sampling.

In general, the set of all states $\underline{x}(0)$ that can be taken into the origin in N sampling periods or less, with $|u(j)| \leqslant 1$, j = 1, 2, ..., N, is given by the set Γ_N:

$$\Gamma_N = \left(\underline{x}(0) \,\middle|\, \underline{x}(0) = \sum_{j=1}^{N} u(j)\underline{r}_j \; ; \; |u(j)| \leqslant 1, \; j = 1, 2, \ldots, N \right). \tag{A-26}$$

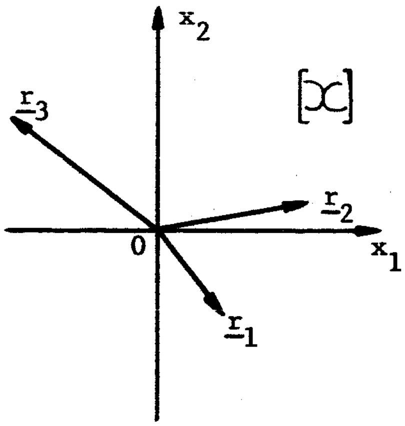
Figure 82 shows the method of generating the set Γ_3 for a second order system.

The following properties of Γ_N can be shown (14, 23, 24, 25):

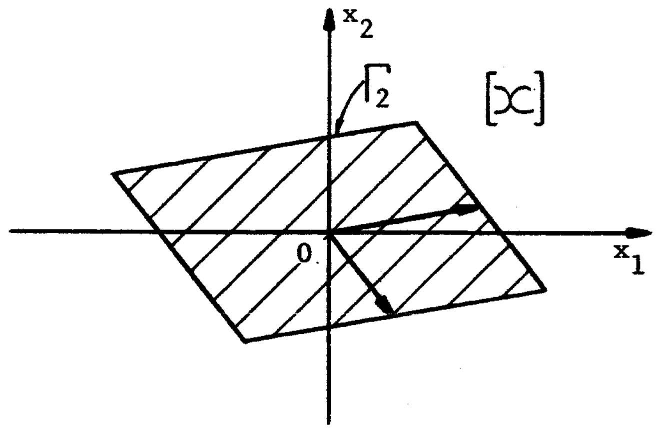
1. Γ_N is a convex set and contains the origin as an interior point.
2. Γ_N is symmetric with respect to the origin.
3. Γ_i is a proper subset of Γ_j for $j > i$.
4. For $T > 0$,
   a. $\lim_{N \to \infty} \Gamma_N = \mathcal{X}$ if and only if $\text{Re.}[\lambda_i] \leqslant 0$, i = 1, 2, ..., n.
   b. $\lim_{N \to \infty} \Gamma_N = \mathcal{X}$ if and only if $|\lambda_i^*| \leqslant 1$, i = 1, 2, ..., n.

Here $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A (the poles of the continuous plant), $\lambda^*_1, \lambda^*_2, \ldots, \lambda^*_n$ are the eigenvalues of G(T) and the system, continuous and discrete, is completely controllable.

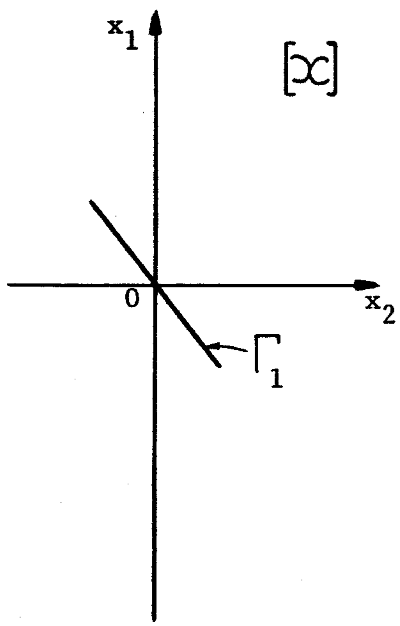
If an initial state $\underline{x}(0)$ is in the set Γ_N then the state may be taken to the origin in N sampling periods or less with an amplitude



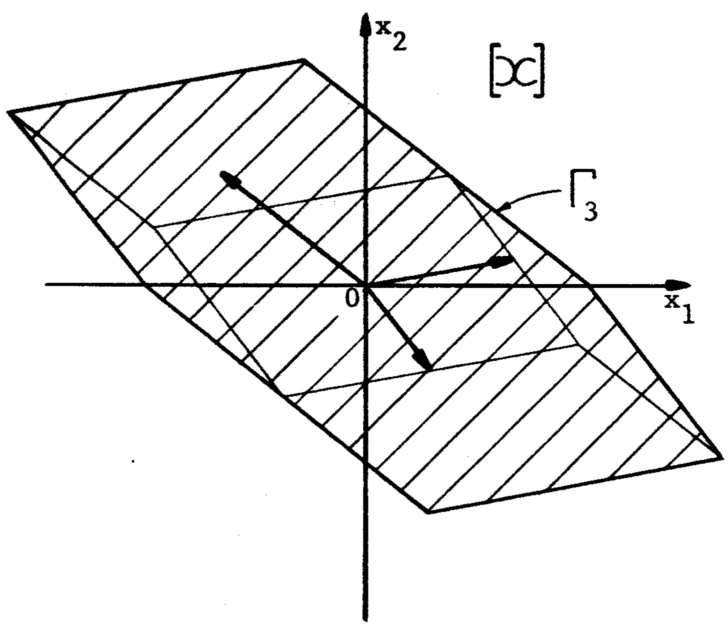
a. The canonical vectors



c. The set Γ_2



b. The set Γ_1



d. The set Γ_3

Figure 82. The generation of Γ_1, Γ_2 and Γ_3 in $\mathcal{X}$-space.

limited input sequence. Conversely, if for a given N, $\underline{x}(0)$ is not Γ_N there is no solution to the (deadbeat regulator) problem of making $\underline{x}(N) = 0$.

The Invariant Vectors

The canonical vectors $\underline{r}_j$ are dependent on the coordinate system of the state space; i.e., they depend on the particular choice of the state variables $x_1, \ldots, x_n$. Since the state of the plant must be referred to some coordinate system, the regulator problem would seem to depend on the choice of state variables. On the other hand, the optimal input sequence must remain the same no matter what coordinate system is chosen to represent the state of the plant. If the initial state is described not by the canonical vectors, but by the invariant vectors discussed below, the regulator problem formulation becomes independent of the choice of state variables. Let the n x n nonsingular matrix R be defined as

$$R = \left[\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_n\right] , \qquad \text{(A-27)}$$

and define the invariant vectors $\underline{h}_j$, j = 1, 2, ..., as

$$\underline{h}_j = R^{-1}\underline{r}_j \quad . \qquad \text{(A-28)}$$

It will be shown that the invariant vectors $\underline{h}_j$, j = 1, 2, ..., are dependent only on the poles of the plant; i.e., the characteristic equation of the matrix A. It will, therefore, follow immediately that:

1. The invariant vectors are independent of the coordinate system of the state space.

2. The invariant vectors are independent of the zeroes of the transfer function.

It is now demonstrated that the invariant vectors are dependent only on the poles of the plant.

Consider a plant with distinct poles,

$s = -\lambda_i$, $i = 1, 2, \ldots, n$,

and represented by the transfer function,

$$\frac{C(s)}{U(s)} = \frac{\prod_{i=1}^{m} (s + z_i)}{\prod_{i=1}^{n} (s + \lambda_i)} \quad , \; m < n \quad . \tag{A-29}$$

Since the poles are distinct, the transfer function can be expanded into partial fractions giving,

$$C(s) = \sum_{i=1}^{n} X_i(s) = \left[\sum_{i=1}^{n} d_i \Big/ (s + \lambda_i) \right] U(s) \quad , \tag{A-30}$$

with

$$X_i(s) = \frac{d_i}{s + \lambda_i} \; U(s) \quad . \tag{A-31}$$

For a completely controllable plant, $d_i \neq 0$, $i = 1, 2, \ldots, n$. Choosing x_i as state variables, leads to the state equation,

$$\dot{x}(t) = A \, \underline{x}(t) + \underline{d} \, u(t) \quad , \tag{A-32}$$

where

$$A = \text{diag.}(-\lambda_1, -\lambda_2, \ldots, -\lambda_n) \quad , \tag{A-33}$$

$$\underline{d} = \text{col.}(d_1, d_2, \ldots, d_n) \quad . \tag{A-34}$$

From Equation (A-3) and Equation (A-6),

$$G(T) = \text{diag.}\left[e^{-\lambda_1 T}, e^{-\lambda_2 T}, \ldots, e^{-\lambda_n T} \right] , \tag{A-35}$$

$$\underline{h}(T) = \text{col.}\left[\frac{d_1}{\lambda_1}\left(1 - e^{-\lambda_1 T}\right), \ldots \frac{d_n}{\lambda_n}\left(1 - e^{-\lambda_n T}\right)\right]. \tag{A-36}$$

Let

$$k_i = -\frac{d_i}{\lambda_i}\left(1 - e^{-\lambda_i T}\right), \quad \gamma_i = \lambda_i T, \; i = 1, 2, \ldots, n \; .$$

$$\tag{A-37}$$

Therefore

$$\underline{h}(T) = \text{col.}(-k_1, -k_2, \ldots, -k_n) \quad , \tag{A-38}$$

$$G(T) = \text{diag.}\left[e^{-\gamma_1}, e^{-\gamma_2}, \ldots, e^{-\gamma_n} \right] . \tag{A-39}$$

The canonical vectors given by Equation (A-19) are

$$\underline{r}_j = -G(-jT)\,\underline{h}(T), \; j = 1, 2, \ldots, \tag{A-40}$$

$$= \text{col.}\left[k_1 e^{j \quad 1}, k_2 e^{j \quad 2}, \ldots, k_n e^{j \quad n} \right] . \tag{A-41}$$

Then the matrix R is given by

$$R = \begin{bmatrix} k_1 e^{\gamma_1} & k_1 e^{2\gamma_1} & \cdots\cdots & k_1 e^{n\gamma_1} \\ k_2 e^{\gamma_2} & k_2 e^{2\gamma_2} & \cdots\cdots & k_2 e^{n\gamma_2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ k_n e^{\gamma_n} & k_n e^{2\gamma_n} & \cdots\cdots & k_n e^{n\gamma_n} \end{bmatrix} . \tag{A-42}$$

Let r^*_{ij} be the i,j-th element of R^{-1}, and R_{ij} the cofactor of the i,j-th element of R;

$$r^*_{ij} = \frac{R_{ij}}{|R|} \quad , \tag{A-43}$$

$$= \frac{(-1)^{i+j}}{k_j} \frac{\begin{bmatrix} e^{\gamma_1} & \cdots & e^{(i-1)\gamma_1} & e^{(i+1)\gamma_1} & \cdots & e^{n\gamma_1} \\ \vdots & & \vdots & \vdots & & \vdots \\ e^{\gamma_{j-1}} & & e^{(i-1)\gamma_{j-1}} & e^{(i+1)\gamma_{j-1}} & \cdots & e^{n\gamma_{j-1}} \\ e^{\gamma_{j+1}} & \cdots & e^{(i-1)\gamma_{j+1}} & e^{(i+1)\gamma_{j+1}} & \cdots & e^{n\gamma_{j+1}} \\ \vdots & & \vdots & \vdots & & \vdots \\ e^{\gamma_n} & & e^{(i-1)\gamma_n} & e^{(i+1)\gamma_n} & \cdots & e^{n\gamma_n} \end{bmatrix}}{\begin{bmatrix} e^{\gamma_1} & \cdots & e^{n\gamma_1} \\ \vdots & & \vdots \\ e^{\gamma_n} & \cdots & e^{n\gamma_n} \end{bmatrix}} , \tag{A-44}$$

Let h_{pi} be the i-th component of $\underline{h}_p$ and r_{pj} be the j-th component of $\underline{r}_p$. From Equation (A-28),

$$h_{pi} = \sum_{j=1}^{n} r^*_{ij} \; r_{pj} \quad . \tag{A-45}$$

Let the n x n matrix V be given by

$$V = \begin{bmatrix} e^{\gamma_1} & \cdots\cdots & e^{n\gamma_1} \\ \vdots & & \vdots \\ e^{\gamma_n} & \cdots\cdots & e^{n\gamma_n} \end{bmatrix} , \tag{A-46}$$

then if V_{ij} is the cofactor of the i,j-th element of V, Equation (A-45) gives

$$h_{pi} = \frac{1}{|V|} \sum_{j=1}^{n} \frac{V_{ij} \; k_j e^{p\gamma_j}}{k_j} = \frac{1}{|V|} \sum_{j=1}^{n} V_{ij} e^{p\gamma_j} . \tag{A-47}$$

From Equation (A-47), for j = 1, 2, ...,

$$\underline{h}_j = \begin{bmatrix} e^{\gamma_1} & e^{2\gamma_1} & \cdots & e^{n\gamma_1} \\ e^{\gamma_2} & e^{2\gamma_2} & \cdots & e^{n\gamma_2} \\ \vdots & \vdots & & \vdots \\ e^{\gamma_n} & e^{2\gamma_n} & \cdots & e^{n\gamma_n} \end{bmatrix}^{-1} \begin{bmatrix} e^{j\gamma_1} \\ e^{j\gamma_2} \\ \vdots \\ e^{j\gamma_n} \end{bmatrix} . \tag{A-48}$$

Equation (A-48) is a general formula for obtaining the invariant vectors. It shows that $\underline{h}_j$ is dependent only on $\gamma_1, \ldots, \gamma_n$, which in turn depend only on the poles of the plant and the sampling period. The matrix V in Equation (A-48) is closely related to the Vandermonde matrix (5). If the plant has repeated roots, V has rows which are equal and the inverse of V does not exist. This is simply because the partial fraction expansion of the plant of Equation (A-30) does not exist if the poles are not distinct. However, the invariant vectors can be obtained for repeated roots from Equation (A-48) by first inverting the matrix, thus forming the expression for $\underline{h}_j$, and then taking the limit as the poles move to the same point. An example of this procedure can be seen in the following paragraphs.

For reference purposes the invariant vectors are calculated for first and second order systems.

First order systems. From Equation (A-48) with $\gamma = \lambda T$,

$$\underline{h}_j = e^{(j-1)\gamma}, \quad j = 1, 2, \ldots \quad . \tag{A-49}$$

Second order systems. From Equation (A-48),

$$\underline{h}_j = \frac{1}{e^{\gamma_2} - e^{\gamma_1}} \begin{bmatrix} -e^{(j-1)\gamma_2+\gamma_1} - e^{(j-1)\gamma_1+\gamma_2} \\ e^{(j-1)\gamma_2} - e^{(j-1)\gamma_1} \end{bmatrix} . \tag{A-50}$$

For notational convenience define

$$w_j = \frac{e^{(j-1)\gamma_2} - e^{(j-1)\gamma_1}}{e^{\gamma_2} - e^{\gamma_1}}, \quad j = 1, 2, \ldots \quad . \tag{A-51}$$

By long division there results,

$$w_{j+2} = \sum_{i=0}^{j} e^{(j-i)\gamma_2+i\gamma_1}, \quad j = 1, 2, \ldots \quad , \tag{A-52}$$

with $w_1 = 0$, $w_2 = 1$. Using Equation (A-52), Equation (A-50) gives

$$\underline{h}_{j+2} = \begin{bmatrix} -e^{\gamma_1+\gamma_2} w_{j+1} \\ w_{j+2} \end{bmatrix} , \quad j = 1, 2, \ldots, \tag{A-53}$$

and

$$\underline{h}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} , \quad \underline{h}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} . \tag{A-54}$$

Table I shows the invariant vectors for various second order systems.

In the general n-th order plant, it can be seen from Equation (A-48) that the first n invariant vectors are always unit vectors:

$$\underline{h}_j = \text{col.}(\delta_{j1}, \ldots, \delta_{jj}, \ldots, \delta_{jn}) . \tag{A-55}$$

It is this fact that is so useful when the minimum energy and minimum fuel sequences are calculated.

The Canonical Space

A fundamental equation in the deadbeat regulator is Equation (A-21),

$$\underline{x}(0) = \sum_{j=1}^{N} u(j)\, \underline{r}_j . \tag{A-56}$$

The initial state $\underline{x}(0)$ in $\mathcal{X}$ can be bought to the origin, if and only if it can be represented by Equation (A-56) for some N. If Equation (A-56) is premultiplied by R^{-1},

TABLE I

THE INVARIANT VECTORS FOR VARIOUS SECOND ORDER PLANTS

Denominator of Plant Transfer Function	s^2	$s(s + a)$	$(s + a)^2$	$(s + a - jb)(s + a + jb)$
$h_{j+2\ 1}$	$-j$	$-S(j) = \dfrac{-e^{aT}\left(e^{jaT} - 1\right)}{e^{aT} - 1}$	$-je^{(j+1)aT}$	$-e^{(j+1)aT}\dfrac{\sin jbT}{\sin bT}$
$h_{j+2\ 2}$	$j + 1$	$S(j) + 1$	$(j+1)e^{jaT}$	$e^{jaT}\dfrac{\sin(j+1)bT}{\sin bT}$

$$R^{-1}\underline{x}(0) = \sum_{j=1}^{N} u(j)\,\underline{h}_j \quad . \tag{A-57}$$

Let

$$\underline{c} = R^{-1}\underline{x}(0) \quad . \tag{A-58}$$

Reserving the symbol $\mathcal{X}$ for the original state space of the plant with coordinates $x_1, x_2, \ldots, x_n$, it is convenient to call $R^{-1}\mathcal{X}$ the canonical vector space or just $\mathcal{C}$-space. For any state $\underline{x}$ in $\mathcal{X}$, there is a corresponding $\underline{c}$ in $\mathcal{C}$ given by Equation (A-58). Let $c_1, \ldots, c_n$ be the coordinates of $\mathcal{C}$-space. Then Equation (A-56) in $\mathcal{C}$-space is

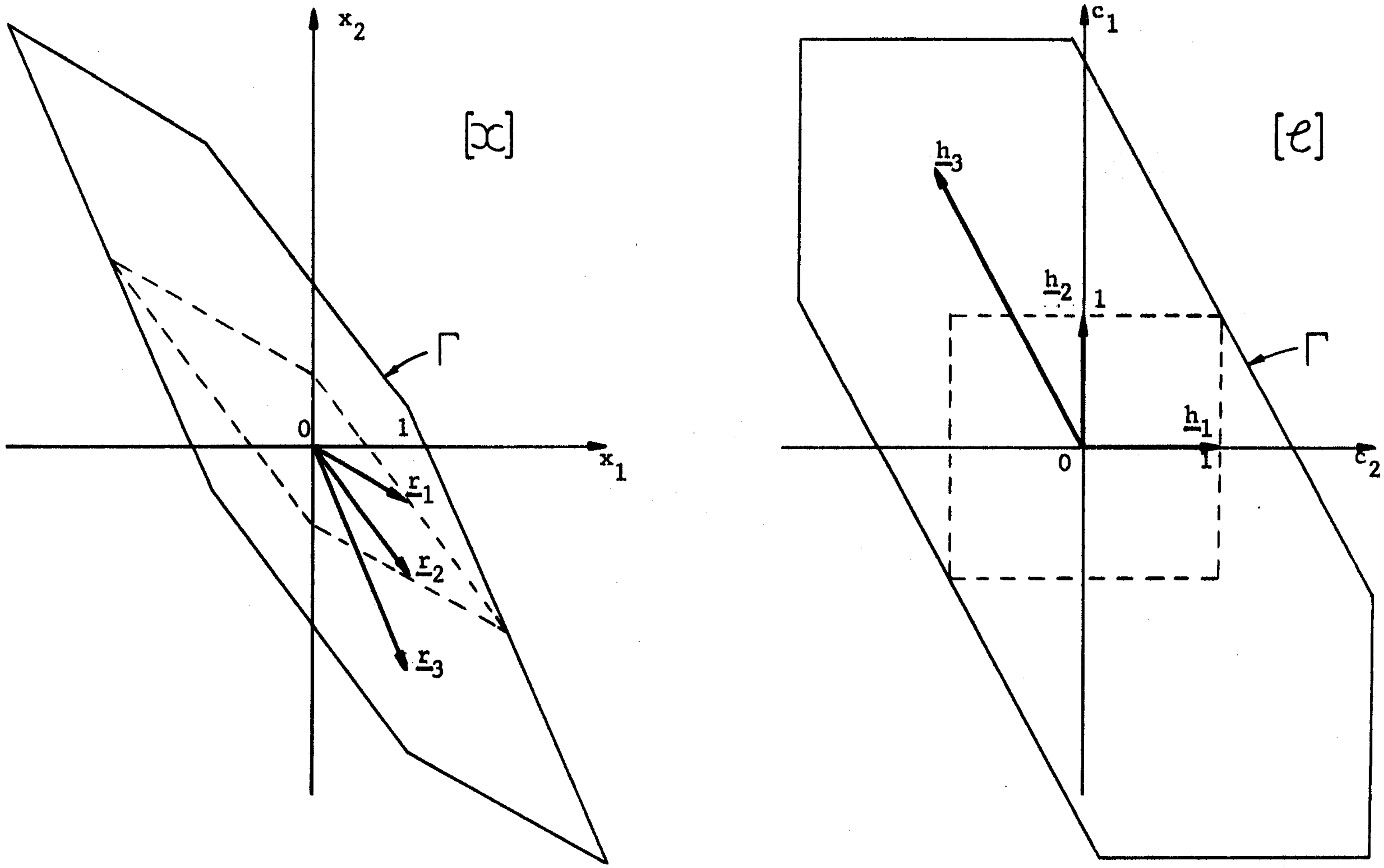
$$\underline{c} = \sum_{j=1}^{N} u(j)\,\underline{h}_j \quad . \tag{A-59}$$

Considering $\underline{c}$ in $\mathcal{C}$ as the state of the plant, the representation of Equation (A-59) is independent of the choice of the state variables, and the properties of the minimum energy and fuel input sequences can be described without reference to any coordinate system since they depend only on the poles of the plant. The matrix R contains all the information on the state space and the zeros of the plant. The formulation of Equation (A-59) is fortunate for another reason: the derivation of the optimal input sequences with this formulation is much simpler compared with the calculations that would be needed if Equation (A-56) were used. This results from the fact that the first n invariant vectors form the columns of the n x n identity matrix (see Chapter II).

As an example of the relation between $\mathcal{X}$ -space and $\mathcal{C}$ -space, the set Γ_N in $\mathcal{C}$ -space is

$$\Gamma_N = \left(\underline{c} \,\middle|\, \underline{c} = \sum_{j=1}^{N} u(j)\underline{h}_j \quad ; \quad \left|u(j)\right| \leqslant 1,\ j = 1,\ 2,\ \ldots,\ N \right) . \tag{A-60}$$

Figure 83 shows Γ_3 for a second order system in both $\mathcal{X}$ and $\mathcal{C}$.



a. Γ_3 in $\mathcal{X}$

b. Γ_3 in $\mathcal{C}$

Figure 83. The set Γ_3 in $\mathcal{X}$ and $\mathcal{C}$ for a second order system.

APPENDIX B

DERIVATION OF EQUATION (3-87) AND THE PROOF OF THEOREM 3

I. THE POINT OF TANGENCY BETWEEN A HYPERELLIPSOID AND A HYPERPLANE

Let the equation of the hyperellipsoid, Equation (3-51), be

$$\underline{\beta}^t \left[I + H^t H\right] \underline{\beta} = \Delta E \quad , \tag{B-1}$$

and let the hyperplane be given by

$$Z \, \underline{\beta} = \underline{z} \quad , \tag{B-2}$$

where the (N-n) x n matrix H is given in Equation (2-17) as

$$H = \left[\underline{h}_{n+1}, \underline{h}_{n+2}, \ldots, \underline{h}_N\right] \quad . \tag{B-3}$$

The matrix Z is assumed to be an r x (N-n) matrix of maximal rank r, $r < (N-n)$, and $\underline{z}$ an r x 1 constant vector. Let the (N-n) x (N-n) matrix Y be defined as

$$Y = \left[I + H^t H\right] \quad . \tag{B-4}$$

The point of tangency between the hyperellipsoid of Equation (B-1) and the (N-n-r)-dimensional hyperplane of Equation (B-2) is the $\underline{\beta}$ which minimizes

$$\underline{\beta}^t Y \underline{\beta} \quad \text{subject to} \quad Z^t \, \underline{\beta} = \underline{z} \quad . \tag{B-5}$$

Since Y is a positive definite matrix, the solution to this problem is given by Equation (2-45) with $\underline{u}^o$ replaced by $\underline{\beta}$, S replaced by Y, C replaced by Z and $\underline{x}(0)$ replaced by $\underline{z}$. Therefore, the point of tangency is given by

$$\underline{\beta} = Y^{-1} Z^t \left[Z Y^{-1} Z^t \right]^{-1} \underline{z} \quad . \tag{B-6}$$

Suppose r = 1, and Z is given by one of the following N possibilities:

$$Z = Z_j = -\underline{p}_j, \quad j = 1, 2, \ldots, n, \tag{B-7}$$

or, for j = n+1, n+2, ..., N,

$$Z = Z_j = \left[\delta_{n+1\ j}, \delta_{n+2\ j}, \ldots, \delta_{N\ j} \right] \tag{B-8}$$

where $\underline{p}_j$ is the j-th row of H, as defined in Equation (3-56) and δ_{ij} is the Kronecker delta. If the 1 x (N-n) matrix Z_j, given by Equations (B-7) and (B-8), is substituted into Equation (B-2), and $\underline{z}$ is replaced by $\delta(j)$, the resulting (N-n-1)-dimensional hyperplane

$$Z_j \underline{\beta} = \delta(j) \tag{B-9}$$

corresponds to the hyperplane W_j if $\delta(j) = -u^o(j) + 1$, and to the hyperplane W_{-j} if $\delta(j) = -u^o(j) - 1$. In this case Equation (B-6) becomes, on replacing $\underline{\beta}$ by $\underline{\beta}_j$,

$$\underline{\beta}_j = \frac{Y^{-1} Z_j^t}{Z_j Y^{-1} Z_j^t} \delta(j) \quad . \tag{B-10}$$

In accordance with Equation (3-47), let

$$\underline{\delta}_j = \begin{bmatrix} \underline{\alpha}_j \\ \\ \underline{\beta}_j \end{bmatrix} \quad , \tag{B-11}$$

where, from Equation (3-54), $\underline{\alpha}_j = -H \underline{\beta}_j$. Therefore, Equations (B-10) and (B-11) give the N-vector $\underline{\delta}_j$. The last N - n members, $\underline{\beta}_j$, come from Equation (B-10) and the first n members from $\underline{\alpha}_j = -H\,\underline{\beta}_j$:

$$\underline{\delta}_j = \frac{1}{Z_j\, Y^{-1}\, Z_j} \begin{bmatrix} -HY^{-1}\, Z_j^t \\ \\ Y^{-1}\, Z_j^t \end{bmatrix} \delta_{(j)} \quad . \qquad \text{(B-12)}$$

II. DERIVATION OF EQUATION (3-87)

Equation (3-87) is

$$\delta_j(i) = \frac{\delta_{ij} - T_{ij}}{1 - T_{jj}}\ \delta_{(j)} \qquad \text{(B-13)}$$

where $\delta_j(i)$ is the i-th member of $\underline{\delta}_j$, $i,j = 1, 2, \ldots, N$, and

$$T_{ij} = u^o(i, \underline{h}_j)\ , \quad i,j = 1, 2, \ldots, N, \qquad \text{(B-14)}$$

is the i-th member of the linear minimum energy input sequence, $\underline{u}^o$, for an initial state $\underline{c} = \underline{h}_j$.

Matrix Identities

Let the n x n matrix X be defined as

$$X = \left[I + HH^t\right] \quad . \qquad \text{(B-15)}$$

Equations (2-25), (2-27), (2-28) and (2-29) are respectively,

$$\underline{b}^o = H^t\, \underline{a}^o \quad , \qquad \text{(B-16)}$$

$$\underline{a}^o = X^{-1}\, \underline{c} \quad , \qquad \text{(B-17)}$$

$$\underline{b}^o = H^t\, X^{-1}\, \underline{c} \quad , \qquad \text{(B-18)}$$

$$\underline{b}^o = Y^{-1}\, H^t\, \underline{c} \quad , \qquad \text{(B-19)}$$

where

$$\underline{u}^o = \begin{bmatrix} \underline{a}^o \\ \\ \underline{b}^o \end{bmatrix} \tag{B-20}$$

is the linear minimum energy input sequence corresponding to the initial state $\underline{c}$. From Equations (B-18) and (B-19),

$$H^t X^{-1} = Y^{-1} H^t \quad . \tag{B-21}$$

Let I denote the identity matrix. Then

$$Y^{-1} Y = I = Y^{-1} \left[I + H^t H\right] = Y^{-1} + Y^{-1} H^t H \quad , \tag{B-22}$$

$$XX^{-1} = I = \left[I + HH^t\right] X^{-1} = X^{-1} + HH^t X^{-1} \quad . \tag{B-23}$$

Postmultiplying and premultiplying Equation (B-21) by H gives

$$H^t X^{-1} H = Y^{-1} H^t H \quad , \tag{B-24}$$

$$HH^t X^{-1} = H Y^{-1} H^t \quad . \tag{B-25}$$

Therefore, Equation (B-22) with Equation (B-24) gives

$$I = Y^{-1} + H^t X^{-1} H \quad , \tag{B-26}$$

and Equation (B-23) with Equation (B-25) gives

$$I = X^{-1} + H Y^{-1} H^t \quad . \tag{B-27}$$

<u>Evaluation of $-H Y^{-1} Z_j^t$, $j = 1, 2, \ldots, N$</u>

From Equation (B-7),

$$\left[Z_1^t, Z_2^t, \ldots, Z_n^t\right] = - \left[\underline{p}_1^t, \underline{p}_2^t, \ldots, \underline{p}_n^t\right] = -H^t \quad . \tag{B-28}$$

Therefore, Equation (B-27) gives

$$-H\, Y^{-1} \left[Z_1^t,\, Z_2^t,\, \ldots,\, Z_n^t \right] = I - X^{-1} \quad . \tag{B-29}$$

Now from Equation (B-8),

$$-H\, Y^{-1} \left[Z_{n+1}^t,\, \ldots,\, Z_N^t \right] = -H\, Y^{-1}\, I \quad , \tag{B-30}$$

so that the transpose of Equation (B-21) gives

$$-H\, Y^{-1} \left[Z_{n+1}^t,\, \ldots,\, Z_N^t \right] = -X^{-1}\, H \quad . \tag{B-31}$$

Therefore, Equations (B-29) and (B-31) give

$$-H\, Y^{-1} \left[Z_1^t,\, \ldots,\, Z_N^t \right] = \left[I - X^{-1},\, -X^{-1}\, H \right] \quad . \tag{B-32}$$

Equation (B-17) gives, with $\underline{c} = \underline{h}_j$, $j = 1, 2, \ldots, N$,

$$\underline{a}^o(\underline{h}_j) = X^{-1}\, \underline{h}_j \quad . \tag{B-33}$$

Therefore, the i,j-th element of $-H\, Y^{-1} \left[Z_1^t,\, \ldots,\, Z_N^t \right]$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, N$, which is the i-th member of $-H\, Y^{-1}\, Z_j^t$, is

$$\delta_{ij} - u^o(i,\, \underline{h}_j) \quad . \tag{B-34}$$

<u>Evaluation of $Y^{-1}\, Z_j^t$, $j = 1, 2, \ldots, N$</u>

$$Y^{-1} \left[Z_1^t,\, \ldots,\, Z_N^t \right] = Y^{-1} \left[-H^t,\, I \right] \tag{B-35}$$

where I is the (N-n) x (N-n) identity matrix. Using Equation (B-21) gives

$$Y^{-1} \left[Z_1^t,\, \ldots,\, Z_N^t \right] = \left[-H^t\, X^{-1},\, Y^{-1} \right] \quad . \tag{B-36}$$

Equation (B-26) gives

$$Y^{-1} = I - H^t\, X^{-1}\, H \quad , \tag{B-37}$$

so that, using Equations (B-16) and (B-33),

$$Y^{-1}\left[Z_1^t, \ldots, Z_N^t\right]$$

$$= \left[-\underline{b}^o(\underline{h}_1), \ldots, -\underline{b}^o(\underline{h}_n),\ I - \left[\underline{b}^o(\underline{h}_{n+1}), \ldots, \underline{b}^o(\underline{h}_N)\right]\right]. \qquad \text{(B-38)}$$

Therefore, the i-th element of $Y^{-1} Z_j^t$, $i = 1, 2, \ldots, N-n$, is

$$\delta_{n+i\ j} = u^o(n + i, \underline{h}_j) \quad . \qquad \text{(B-39)}$$

The matrix of Equation (B-12) is therefore,

$$\frac{1}{Z_j\ Y^{-1}\ Z_j^t} \begin{bmatrix} -H\ Y^{-1}\ Z_j^t \\ \\ Y^{-1}\ Z_j \end{bmatrix} = \frac{1}{Z_j\ Y^{-1}\ Z_j^t}\left[\underline{e}_j - \underline{u}^o(\underline{h}_j)\right] \qquad \text{(B-40)}$$

where $\underline{e}_j = \text{col.}\ \left[\delta_{j1}, \ldots, \delta_{jj}, \ldots, \delta_{jN}\right]$. It therefore remains to evaluate $Z_j\ Y^{-1}\ Z_j^t$. From Equations (B-29) and (B-28),

$$I - X^{-1} = H\ Y^{-1}\ H^t = \begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_n \end{bmatrix} Y^{-1}\left[Z_1^t, Z_2^t, \ldots, Z_n^t\right], \qquad \text{(B-41)}$$

$$= \begin{bmatrix} Z_1\ Y^{-1}\ Z_1^t & \cdots & Z_1\ Y^{-1}\ Z_n^t \\ \vdots & & \vdots \\ Z_n\ Y^{-1}\ Z_1^t & \cdots & Z_n\ Y^{-1}\ Z_n^t \end{bmatrix} . \qquad \text{(B-42}$$

Therefore, from Equations (B-33) and (B-35),

$$Z_j\ Y^{-1}\ Z_j^t = 1 - u^o(j, \underline{h}_j)\ , \quad j = 1, 2, \ldots, n. \tag{B-43}$$

Equation (B-38) gives

$$Z_j\ Y^{-1}\ Z_j^t = 1 - u^o(j, \underline{h}_j)\ , \quad j = n+1, \ldots, N. \tag{B-44}$$

Finally therefore, Equations (B-40), (B-43) and (B-44) give

$$\frac{1}{Z_j\ Y^{-1}\ Z_j^t} \begin{bmatrix} -H\ Y^{-1}\ Z_j^t \\ \\ Y^{-1}\ Z_j \end{bmatrix} = \frac{\underline{e}_j - \underline{u}^o(\underline{h}_j)}{1 - u^o(j, \underline{h}_j)}\ . \tag{B-45}$$

Therefore, with Equation (B-14), there results from Equations (B-12) and (B-45),

$$\delta_j(i) = \frac{\delta_{ij} - T_{ij}}{1 - T_{jj}}\ \delta(j)\ , \tag{B-46}$$

which is the desired form of Equation (B-13).

Useful Properties of T_{ij}

Since $\left[I + HH^t\right]^{-1}$ is a positive definite symmetric matrix, a matrix D can be found such that

$$D^t\ D = \left[I + HH^t\right]^{-1}\ . \tag{B-47}$$

Define the n-vector $\underline{n}_j$ as

$$\underline{n}_j = D\ \underline{h}_j, \quad j = 1, 2, \ldots, N\ . \tag{B-48}$$

Then

$$\underline{n}_i^t\ \underline{n}_j = \underline{h}_i^t\ D^t\ D\ \underline{h}_j\ , \tag{B-49}$$

which from Equation (B-16) gives

$$\underline{n}_i^t \, \underline{n}_j = T_{ij} \quad . \tag{B-50}$$

Consider an initial state $\underline{c} = \underline{h}_j$. One possible input sequence which brings $\underline{c}$ to the origin is $u(j) = -1$, $u(i) = 0$, $i \neq j$. This input sequence has an energy $E = 1$. The minimum energy consumption is, from Equation (2-39),

$$E^o = \underline{h}_j^t \, D^t \, D \, \underline{h}_j = \underline{n}_j^t \, \underline{n}_j = T_{jj} \quad , \tag{B-51}$$

but $0 < E^o \leqslant E = 1$. Therefore,

$$0 < T_{jj} \leqslant 1, \quad j = 1, 2, \ldots, N \, . \tag{B-52}$$

If $\underline{c}$ is in Γ_N, and if for any j, $j = 1, 2, \ldots, N$, $u^o(j; \underline{c}) > 1$, then $0 < E^o < 1$, and therefore

$$0 < T_{jj} < 1, \quad j = 1, 2, \ldots, N. \tag{B-53}$$

From the Schwartz inequality,

$$\left| \underline{n}_i^t \, \underline{n}_j \right| \leqslant \sqrt{\left[\underline{n}_i^t \, \underline{n}_i \right] \left[\underline{n}_j^t \, \underline{n}_j \right]} \leqslant 1 \quad . \tag{B-54}$$

Therefore,

$$\left| T_{ij} \right| \leqslant 1, \quad i,j = 1, 2, \ldots, N \quad . \tag{B-55}$$

Finally, since $\underline{n}_i^t \, \underline{n}_j = \underline{n}_j^t \, \underline{n}_i$,

$$T_{ij} = T_{ji}, \quad i,j = 1, 2, \ldots, N. \tag{B-56}$$

This result is helpful when applying Theorem 3, since only the values of T_{ij}, i,j in J, $i \geqslant j$, need be to be used. Note further, that T_{ij}, $i,j \leqslant n$, can be found as the elements of the matrix $\left[I + HH^t \right]^{-1}$, and T_{ij}, $i,j \geqslant n+1$, correspond to the elements of the matrix $\left[I + H^t H \right]^{-1}$.

III. PROOF OF THEOREM 3

Statement of Theorem 3

Theorem 3 applies to initial states $\underline{c}$ in the set Γ_N but not in M_N. The set of all j for which $|u^o(j)| > 1$ defines the set J. Without loss of generality assume that $u^o(1) > 1$. Then from Equation (3-87)

$$\delta_j(i) = \frac{\delta_{ij} - T_{ij}}{1 - T_{jj}} \quad \delta(j) \tag{B-57}$$

where $\delta(j) = \text{sgn. } u^o(j) - u^o(j)$ from Equation (3-91). The N-vector $\underline{\delta}_j$ is given by

$$\underline{\delta}_j = \begin{bmatrix} \delta_j(1) \\ \cdot \\ \cdot \\ \cdot \\ \delta_j(N) \end{bmatrix} . \tag{B-58}$$

Theorem 3 states that

$$u^e(1) = 1 \tag{B-59}$$

if, for all j in J,

$$u^o(1) + \delta_j(1) \geqslant 1 \tag{B-60}$$

and $\underline{c}' = \underline{c} - \underline{h}_1$ is in the set

$$\Gamma'_{N-1} = \left(\underline{c}' \middle| \; \underline{c}' = \sum_{j=2}^{N} u(j) \, \underline{h}_j ; \quad |u(j)| \leqslant 1, \; j = 2, 3, \ldots, N \right). \tag{B-61}$$

Proof of Theorem 3

This proof is based on the work of Stubberud and Swiger (36, page 405). The correction $\underline{\beta}^e$ must lie on the boundary of U_{N-n}, so that for at least one j in J, $\delta^e(j) = \text{sgn. } u^o(j) - u^o(j)$. From Equation (B-50), the vector $\underline{\delta}_j + \mu\left[\underline{\delta}^e - \underline{\delta}_j\right]$ has a first component,

$$\delta(1) = \delta_j(1) + \mu\left[\delta^e(1) - \delta_j(1)\right] \tag{B-62}$$

equal to $\delta(1) = 1 - u^o(1)$ for some value of μ, and has an energy correction cost less than or equal to that of $\underline{\delta}^e$. But since $\underline{c}'$ is in Γ'_{N-1}, this value of μ gives the corresponding vector $\underline{\beta}^e$ in U_{N-n}. Therefore $\delta^e(1) = 1 - u^o$, and $u^e(1) = 1$.

APPENDIX C

SYMBOLISM

Only frequently recurring symbols are included below: the meaning of any other notation should be apparent from the accompanying text. The underscoring of a symbol represents vector notation; the subscripted symbol without underscoring represents the components of the corresponding vector. The letters i, j, k, m, and p always represent either an integer or zero, while μ and α are used as arbitrary constants. Capital roman letters usually represent matrices or particular sets: the exception being when A, B, C, $\bar{A}$, $\bar{B}$ and $\bar{C}$ are used to denote general regions of interest in the figures. The symbol 0 is used to represent the corresponding scalar, vector or matrix, the particular use in the text is apparent.

I. LIST OF SYMBOLS

SYMBOL	MEANING
$\underline{a}$	The n-vector, consisting of the first n members of the input sequence, with components $a_1, a_2, \ldots, a_n$.
$\underline{a}^o$	The n-vector of the first n members of the linear energy optimum input sequence, with components $a_1^o, a_2^o, \ldots, a_n^o$.
$\underline{b}$	The (N-n)-vector of the last (N-n) input sequence members, with components $b_1^o, b_2^o, \ldots, b_{N-n}^o$.

SYMBOL	MEANING
$\underline{c}$	The state of the plant in $\mathcal{C}$ -space, with components c_1, c_2, ..., c_n.
$c_k(j)$	Points in a one-dimensional $\mathcal{C}$ -space, defined by Equation (5-76), page 202.
$\underline{d}$	The forcing vector of the continuous plant, with components d_1, d_2, ..., d_n.
$\underline{e}_j$	The 1 x N matrix, $\left[\delta_{j1}, \delta_{j2}, \ldots, \delta_{jj}, \ldots, \delta_{jN}\right]$.
e	The base for natural logarithms.
$\underline{f}$	A vector feedback function.
f	A particular fuel consumption, or a scalar feedback function.
h(T)	Forcing vector of the discrete plant.
$\underline{h}_j$	The j-th invariant vector, defined in Equation (A-28), page 245, with components h_{j1}, h_{j2}, ..., h_{jn}.
ℓ_j	The length (Euclidean norm) of the j-th invariant vector, defined in Equation (2-46), page 24.
n	The order of the plant.
p	Number of members of the set k.
$\underline{p}_j$	A 1 x N matrix, the j-th row of the matrix H, defined in Equation (3-56), page 94.
$\underline{r}_j$	The j-th canonical vector, defined in Equation (A-19), page 239, with components r_{j1}, r_{j2}, ..., r_{jn}.
s	Laplace transform variable.

SYMBOL	MEANING
t	The transpose of a matrix or the time.
$\underline{u}$	The N-vector representing the discrete plant input sequence.
$\underline{u}^o$	The linear minimum energy input sequence.
$\underline{u}^e$	The constrained minimum energy input sequence.
$\underline{u}^f$	The constrained minimum fuel input sequence.
$u(t)$	The plant input.
$\underline{x}(t)$	The plant state vector, an n-vector, with components $x_1(t), x_2(t), \ldots, x_n(t)$.
$\underline{x}(k)$	The state vector at time $t = kT$.
$u^o(i,\underline{h}_j)$	The i-th member of $\underline{u}^o$ when the initial state is $\underline{c} = \underline{h}_j$.
A	The set, in $\mathcal{A}$-space, of all $\underline{a}$ whose members satisfy Equation (4-21), page 142.
A'	The map of A in $\mathcal{B}$-space.
A_n	The set in α-space, composed of all the $\underline{\alpha}$ that satisfy Equation (3-52), page 92.
A_n'	The map of A_n in β-space.
B	A set, in $\mathcal{B}$-space, of all $\underline{b}$ whose members satisfy Equation (4-22), page 142.
B_{N-n}	A set, in β-space, composed of all the $\underline{\beta}$ that satisfy Equation (3-53), page 92.
C	The n x N matrix of the first N canonical vectors, $\underline{r}_j$.
$C(s)$	Laplace transform of plant output.

SYMBOL	MEANING
$C(\underline{+}i,\underline{+}j)$	The cone in $\mathcal{C}$-space generated by the line $L(\underline{+}i,\underline{+}j)$, see page 31.
$C_s(\underline{+}i,\underline{+}j)$	The cone in $\mathcal{C}$-space generated by the line $L_s(\underline{+}i,\underline{+}j)$, see page 40.
E	The energy cost, defined in Equation (1-6), page 9.
E^o	The energy cost associated with $\underline{u}^o$.
E^e	The energy cost associated with $\underline{u}^e$.
F	The fuel cost, defined in Equation (1-7), page 9.
F_A	The fuel cost associated with $\underline{a}$, defined by Equation (4-25), page 143.
F_B	The fuel cost associated with $\underline{b}$, defined by Equation (4-26), page 143.
F_N	The set of all initial states, $\underline{c}$, whose linear minimum fuel sequence satisfies the saturation constraint.
$G(T)$	The n x n state transition matrix, see page 237.
$G_p(s)$	The transfer function of the continuous plant.
H	The n x (N-n) matrix whose columns are the last N-n invariant vectors. The i,j-th element of H is the i-th component of the (n + j)-th invariant vector, $\underline{h}_j$, h_{ji}.
I	The identity matrix of the same order as the matrix with which it may be associated.
J	The set of integers j for which $\left\|u^o(j)\right\| > 1$.
K	The set of integers k, for which $\underline{h}_k$ lies on $\partial S_N(f)$.

SYMBOL	MEANING
L_N	The set, in $\mathcal{A}^o$-space, of all $\underline{a}^o$ satisfying Equation (3-31), page 84.
$L(i,j)$	The line $\mu\, \underline{h}_i + (1 - \mu)\, \underline{h}_j$, $0 \leq \mu \leq 1$, joining any pair of points $\underline{h}_i$ and $\underline{h}_j$ in a two-dimensional $\mathcal{C}$-space.
$L_s(i,j)$	The line $\mu\, \underline{h}_i + (1 - \mu)\, \underline{h}_j$, $0 \leq \mu \leq 1$, i,j in K, in a two-dimensional $\mathcal{C}$-space, and which belongs to $\partial S_N(1)$.
M_N	The set, in $\mathcal{C}$-space, of all initial states whose linear minimum energy input sequence satisfies the saturation constraints, defined by Equation (3-26), page 80.
N	The number of sampling periods for the deadbeat control.
Q	The n x (N-n) matrix whose columns are the last (N-n) canonical vectors defined in Equation (2-9), page 19.
Q_k	A polygonal region in $\mathcal{C}$-space, which, for k = N, N-1, ..., 1, defines the feedback required for the solution of the constrained minimum fuel problem.
R	The n x n matrix whose columns are the first n canonical vectors, $\underline{r}_j$, defined in Equation (2-8), page 19.
S(j)	Notation convenient in describing the invariant vectors for second order plants with integration, defined in Table I, page 252.
T	The sampling period of the discrete regulator system.
T_{ij}	An alternate notation for $u^o(i,\underline{h}_j)$, i,j = 1, 2, ..., N.

SYMBOL	MEANING
$S_k(f)$	The set, in $\mathcal{C}$-space, of all initial states which can be taken to the origin in k sampling periods or less with a fuel consumption $F \leqslant f$.
U	The set, in $\mathcal{B}$-space, defined by the intersection of the sets B and A'.
U_{N-n}	The set, in β-space, defined by the intersection of the sets B_{N-n} and A'_n.
U(s)	The Laplace transform of u(t).
$W_{\pm j}$	The (N-n)-dimensional hyperplanes defined by Equations (3-73) and (3-74), page 101.
$W_{\pm(n+j)}$	The (N-n)-dimensional hyperplanes defined by Equations (3-71) and (3-72), page 100.
X	The n x n matrix $\left[I + HH^t\right]$.
Y	The (N-n) x (N-n) matrix $\left[I + H^tH\right]$.
Z_j	The 1 x (N-n) row vector defined by Equations (B-7) and (B-8), page 256.
$\underline{\alpha}$	The n x 1 correction vector, formed as the first n components of $\underline{\delta}$.
$\underline{\alpha}^e$	The optimum correction, which, when added to $\underline{a}^o$, gives the first n members of $\underline{u}^e$.
$\underline{\beta}$	An (N-n) x 1 correction vector, formed from the last (N-n) components of $\underline{\delta}$.
$\underline{\beta}^e$	An optimum correction, which, when added to $\underline{b}^o$, gives the last (N-n) members of $\underline{u}^e$.

SYMBOL	MEANING
$\underline{\beta}^f$	An optimum correction, which, when added to $\underline{b}^o$, gives the last (N-n) members of $\underline{u}^f$.
$\underline{\beta}_{\pm j}$	The point of tangency of the hyperellipsoid with the hyperplane $W_{\pm j}$.
γ_i	Scalar constant, given by $\gamma_i = \lambda_i T$.
γ	Scalar constant, given by $\gamma = \lambda T$.
$\underline{\delta}$	An N x 1 correction vector with components $\delta(1), \ldots, \delta(N)$.
$\underline{\delta}^e$	The optimum correction for the constrained minimum energy input sequence.
$\underline{\delta}^f$	The optimum correction for the constrained minimum fuel input sequence.
$\underline{\delta}_j$	The correction corresponding to the point of tangency of the hyperellipsoid with the hyperplane $\delta(j)$ = constant, with components $\delta_j(1), \delta_j(2), \ldots, \delta_j(N)$.
δ_{ij}	The Kronecker delta, defined by Equation (2-82), page 42.
$\partial S_N(f)$	The boundary of the set $S_N(f)$.
θ_i	The N-n eigenvalues of the matrix $\left[I + H^t H\right]$.
$-\lambda_i, -\lambda$	The poles of the continuous plant.
τ	A dummy variable.
$\underline{\omega}_i$	The eigenvectors of the matrix $\left[I + H^t H\right]$.
Γ_N	The set of all initial states that can be taken to the origin with an amplitude constrained input sequence.

SYMBOL	MEANING
Γ_N cont.	In $\mathcal{C}$-space, the set is defined by either Equation (3-27), page 80, or Equation (A-60), page 254. In $\mathcal{X}$-space the set is defined by Equation (A-26), page 243.
Γ_k'	The set of all states that can be taken to the origin of $\mathcal{C}$-space when only k (not necessarily the first k) of the invariant vectors are available to represent the state.
ΔE	The energy cost associated with a correction $\underline{\delta}$, see page 89.
Φ	A line in two-dimensional $\mathcal{C}$-space.
$\mathcal{A}$-space	The n-dimensional space with coordinates $a_1, a_2, \ldots, a_n$.
$\mathcal{A}^o$-space	The n-dimensional space with coordinates $a_1^o, a_2^o, \ldots, a_n^o$.
$\mathcal{B}$-space	The (N-n)-dimensional space with coordinates $b_1, b_2, \ldots, b_{N-n}$.
$\mathcal{C}$-space	The n-dimensional space with coordinates $c_1, c_2, \ldots, c_n$.
α-space	The n-dimensional space with coordinates $\alpha_1, \alpha_2, \ldots, \alpha_n$.
β-space	The (N-n)-dimensional space with coordinates $\beta_1, \beta_2, \ldots, \beta_{N-n}$.
$\mathcal{X}$-space	The n-dimensional state space with coordinates $x_1, x_2, \ldots, x_n$.

II. MATHEMATICAL NOTATION

SYMBOL	MEANING
$\|A\|$	The determinant of a square matrix A.
$\|\alpha\|$	The magnitude of a scalar α.
sat. α	$= \begin{cases} 1 & \text{if } \alpha > 1 \\ \alpha & \text{if } \|\alpha\| < 1 \\ -1 & \text{if } \alpha < -1 \end{cases}$.
sgn. α	$= \begin{cases} 1 & \text{if } \alpha \geqslant 0 \\ -1 & \text{if } \alpha < 0 \end{cases}$.
$A = \left(\underline{x} \mid \underline{x} = B;\ C \right)$	The set A is defined as the set of all $\underline{x}$ given by $\underline{x} = B$, subject to C.